Chapter 18

Fields and polynomials

Introduction

In this chapter we embark on a systematic study of fields. As you know from Book C, fields are a special sort of ring – loosely, one in which division is possible by every element but zero. We have already met several examples of fields. In addition to the familiar number fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, we worked with the finite field of p elements, where p is a prime, and with fields such as $\mathbb{Q}[\sqrt{2}]$ that contain $\mathbb{Q}$ and are contained in $\mathbb{C}$.

In Book C, you saw the relevance of polynomials over a field to number theory. Polynomials over a field can also represent geometric curves, so another application of fields is that they provide a suitable context to treat certain geometric questions by algebraic methods – this is what is done in a central branch of mathematics called algebraic geometry. Later in this book, in Chapter 20, we will discuss a different and much more recent application of field theory, namely certain methods for coding messages that are essential for secure online information exchange.

We start the chapter by discussing field isomorphisms, and then we look at further constructions of fields. In Chapter 17 we saw that the quotient of a ring by a maximal ideal is in fact a field. A particular case of this construction is obtained by taking the quotient of the polynomial ring $F[x]$ over a field F by a maximal ideal $\langle p(x)\rangle$, where $p(x)$ is an irreducible polynomial. For example, the polynomial $p(x) = x^2 - 2$ in $\mathbb{Q}[x]$ gives rise in this way to a field that is isomorphic to $\mathbb{Q}[\sqrt{2}]$. We find an interesting contrast between the case where $F = \mathbb{Q}$ and where F is a finite field.

We then introduce the *characteristic* of a field, which is the least positive number m such that

$$\underbrace{1+1+1+\cdots+1}_{m \text{ times}} = 0.$$

When this number exists, it must be a prime p, and we then say that the field has characteristic p. When no such number exists, we say that the field has characteristic zero. So, for example, a finite field with p elements has characteristic p, and the fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ have characteristic zero.

Section 2 looks at field extensions. Whenever a field E contains an isomorphic copy of another field F, we say that E is an *extension field* of F. For example, when we quotient a polynomial ring $F[x]$ over a field F by maximal ideal M in $F[x]$, the quotient $F[x]/M$ contains (a copy of) F as a subfield, so $F[x]/M$ is an extension of F. It turns out that any extension of a field F is also a vector space over F. Since vector spaces over arbitrary fields may not be familiar to you, we define them and list their basic properties.

If E is a field extension of F and α is an element of E that is a root of a non-zero polynomial in $F[x]$, then we say that α is *algebraic* over F. Otherwise, we say that α is *transcendental* over F. Familiar examples of elements of $\mathbb{R}$ that are transcendental over $\mathbb{Q}$ are the numbers e and π. We conclude Section 2 with a description of the different properties of

algebraic extensions (those where every element is algebraic over the base field) and transcendental ones.

Section 3 looks at finite fields in some detail. We show that the multiplicative group of such fields is cyclic and we practise calculations in finite fields.

In Section 4 we look at field extensions that contain all the roots of a given polynomial $p(x)$, and introduce the notion of a *splitting field* for $p(x)$, that is, the smallest extension of the field that contains all the roots of $p(x)$. The chapter ends with a classification of finite fields, achieved by showing that for any prime power p^n there is a finite field of order p^n, and then that any two such fields are isomorphic.

Chapter guide

The second section is the most substantial and we expect it to require about twice the study time as the other sections. Sections 1, 3 and 4 have similar contents and we expect them to require similar study times. Many of the exercises involve working modulo p for a prime p.

1 Preliminaries

In this section we introduce two notions that are fundamental to our work on fields, that is, isomorphisms between fields and the characteristic of a field.

1.1 Isomorphisms and homomorphisms between fields

In Definition 1.4 of Chapter 17, we introduced ring homomorphisms. Since fields are a special type of ring, it is natural to ask whether the existence of a ring isomorphism between two fields also preserves the field structure – in other words, whether the isomorphism preserves inverses.

The next proposition shows that this is in fact the case for any ring homomorphism, hence in particular for an isomorphism.

Proposition 1.1

Let F_1 and F_2 be fields and suppose that $\theta : F_1 \to F_2$ is a ring homomorphism. If $a \in F_1$ with $a \neq 0$, then

$$\theta(a^{-1}) = \theta(a)^{-1}.$$

Proof Since θ is a ring homomorphism,

$$\theta(a)\theta(a^{-1}) = \theta(aa^{-1}) = \theta(1_{F_1}) = 1_{F_2}.$$

Thus $\theta(a^{-1}) = \theta(a)^{-1}$ as required. ■

Proposition 1.1 allows us to define two fields to be isomorphic precisely when they are isomorphic as rings.

Definition 1.2 *Isomorphism of fields*

Let F_1 and F_2 be fields. We say that F_1 and F_2 are **isomorphic** if, and only if, there is a ring isomorphism $\theta : F_1 \to F_2$. We then also say that θ is a **field isomorphism**.

We illustrate this definition by working out a particular isomorphism of two infinite fields.

Example 1.3 *A field isomorphism*

In Example 1.7(b) of Chapter 17 we met the field of rational functions $\mathbb{Q}(x)$, which can be constructed as the field of fractions of the polynomial ring $\mathbb{Q}[x]$.

Let $r(x) = \dfrac{s(x)}{t(x)}$ be a rational function over $\mathbb{Q}$, that is, an element of $\mathbb{Q}(x)$ and let $\alpha \in \mathbb{C}$ be such that $t(\alpha) \neq 0$. Denote by $r(\alpha)$ the evaluation of the rational function $r(x)$, that is, the value $\dfrac{s(\alpha)}{t(\alpha)}$.

Now let us define two sets A and B as follows:

$$A = \{r(e) : r(x) \text{ a rational function over } \mathbb{Q}\}, \quad \text{and}$$
$$B = \{r(\pi) : r(x) \text{ a rational function over } \mathbb{Q}\},$$

where e denotes the base of the natural logarithm.

It is known that neither e nor π are roots of a polynomial in $\mathbb{Q}[x]$, so both these sets are well-defined subsets of $\mathbb{R}$. A little more detail on this will be given in Example 2.18. Since rational functions are closed under addition, multiplication, subtraction and (non-zero) division, and since both sets contain 1, it follows that A and B are both subfields of $\mathbb{R}$.

Let $\theta : A \to B$ be defined by

$$\theta : r(e) \mapsto r(\pi),$$

so that, for example, $\theta(1 + e) = 1 + \pi$. Then θ is a field isomorphism. We will show this in Corollary 2.32.

In the next exercise, you are asked to prove a result on ring isomorphisms that we will use repeatedly.

Exercise 1.1

Let F be a field and R be a ring. Suppose that $f\colon F \to R$ is a non-zero ring homomorphism, that is, such that $\mathrm{Im}(f) \neq \{0\}$. Show that $\mathrm{Im}(f)$ is a field isomorphic to F.

Maps where the image of the domain is isomorphic to the domain are known as **embeddings**. Exercise 1.1 proves the following proposition.

Proposition 1.4

Let F and K be fields and suppose that $f\colon F \to K$ is a non-zero ring homomorphism. Then f is an embedding, that is $\mathrm{Im}\, f \cong F$.

We next prove a proposition about quotient fields of polynomial rings that will be repeatedly used in this chapter.

Proposition 1.5

Let F be a field and let $p(x)$ be an irreducible polynomial in $F[x]$. Then $E = F[x]/\langle p(x)\rangle$ is a field.

Proof $F[x]$ is a Euclidean domain (by Example 3.2 of Book C, Chapter 12) and so, by Corollary 2.11 of Chapter 17, $F[x]$ is a PID. Thus, by Theorem 4.22 of Chapter 17, as the polynomial $p(x)$ is irreducible over $F[x]$, the ideal $\langle p(x)\rangle$ is maximal. We can deduce from Theorem 4.23 of Chapter 17 that E is a field. ■

In the next example you will see two particular instances of fields of the form described in Proposition 1.5. As you will see while reading this chapter, this type of construction is a rich source of fields. In fact, in the last section you will learn that all finite fields arise as quotients of this kind.

Example 1.6 *Fields as quotients of polynomial rings*

(a) Let $K = \mathbb{Q}[x]/\langle x^2 - 2\rangle$. Since $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$, K is a field by Proposition 1.5. Let $\theta : \mathbb{Q}[x] \to K$ be the natural homomorphism.

Let $\alpha = \theta(x) = x + \langle x^2 - 2\rangle$. Our aim is to write the elements of K in terms of $\mathbb{Q}$ and α.

For any $f(x) \in \mathbb{Q}[x]$ we use the Division Algorithm to write

$$\begin{aligned} f(x) &= q(x)(x^2 - 2) + r(x) \text{ where } \deg r(x) < 2 \\ &= q(x)(x^2 - 2) + a + bx \text{ for some } a, b \in \mathbb{Q}. \end{aligned}$$

Rearranging terms gives

$$f(x) - (a + bx) = q(x)(x^2 - 2) \in \langle x^2 - 2 \rangle$$

or, equivalently by Proposition 4.2 of Chapter 17,

$$f(x) + \langle x^2 - 2 \rangle = a + bx + \langle x^2 - 2 \rangle = a + b\alpha.$$

Hence every element $f(x) + \langle x^2 - 2 \rangle$ of K can be written uniquely as $a + b\alpha$ with $a, b \in \mathbb{Q}$. Consequently, we can regard K as a field whose underlying set is $\{a + b\alpha : a, b \in \mathbb{Q}\}$, and where addition and multiplication are defined as follows.

$$\begin{aligned}(a + b\alpha) + (c + d\alpha) &= (a + c) + (b + d)\alpha \\ (a + b\alpha)(c + d\alpha) &= ac + bd\alpha^2 + (ad + bc)\alpha \\ &= ac + 2bd + (ad + bc)\alpha\end{aligned}$$

(since $\alpha^2 = x^2 + \langle x^2 - 2 \rangle = 2 + x^2 - 2 + \langle x^2 - 2 \rangle = 2 + \langle x^2 - 2 \rangle$)

Note that addition and multiplication of constant polynomials correspond to addition and multiplication in $\mathbb{Q}$, so that there is a one–one homomorphism from $\mathbb{Q}$ to K.

In Book C, Chapter 11, Exercise 1.10, the field $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ was introduced as a subfield of $\mathbb{R}$. Since in $\mathbb{Q}[\sqrt{2}]$ we have

$$(a + b\sqrt{2})(c + d\sqrt{2}) = ac + 2bd + (ad + bc)\sqrt{2},$$

it follows from the preceding calculations that the map $\psi : K \to \mathbb{Q}[\sqrt{2}]$ defined by

$$\psi : a + b\alpha \mapsto a + b\sqrt{2}$$

is an isomorphism.

(b) Let $L = \mathbb{Z}_7[x]/\langle x^2 - 3 \rangle$. Since 3 is a quadratic non-residue over 7, $x^2 - 3$ is irreducible over $\mathbb{Z}_7[x]$ and, by Proposition 1.5, L is a field. Let $\theta : \mathbb{Z}_7[x] \to L$ be the natural homomorphism.

Let $\alpha = \theta(x) = x + \langle x^2 - 3 \rangle$. As in part (a) we aim to show that the elements of L can be written in terms of $\mathbb{Z}_7$ and α.

If $f(x) \in \mathbb{Z}_7[x]$, we use the Division Algorithm to write

$$\begin{aligned}f(x) &= q(x)(x^2 - 3) + r(x) \text{ where } \deg r(x) < 2 \\ &= q(x)(x^2 - 3) + a + bx \text{ for some } a, b \in \mathbb{Z}_7.\end{aligned}$$

Thus, we have

$$f(x) - (a + bx) = q(x)(x^2 - 3) \in \langle x^2 - 3 \rangle,$$

which gives

$$f(x) + \langle x^2 - 3 \rangle = a + bx + \langle x^2 - 3 \rangle = a + b\alpha.$$

It follows that every element of L can be written uniquely as $a + bx + \langle x^2 - 3\rangle = a + b\alpha$, where $a, b \in \mathbb{Z}_7$. Hence we can regard L as a field with underlying set $\{a + b\alpha : a, b \in \mathbb{Z}_7\}$ and with addition and multiplication defined by

$$\begin{aligned}(a + b\alpha) + (c + d\alpha) &= (a + c) + (b + d)\alpha \pmod 7\\ (a + b\alpha)(c + d\alpha) &= ac + bd\alpha^2 + (ad + bc)\alpha\\ &= ac + 3bd + (ad + bc)\alpha \pmod 7\end{aligned}$$

since $\alpha^2 = x^2 + \langle x^2 - 3\rangle = 3 + (x^2 - 3) + \langle x^2 - 3\rangle = 3 + \langle x^2 - 3\rangle$.

The next proposition generalises Examples 1.6(a) and (b) to any field F and irreducible polynomial in $F[x]$. In the proof, division in the ring $F[x]$ can be carried out using the Division Algorithm given in Book C, Chapter 11, Theorem 3.1.

Proposition 1.7

Let $p(x) = p_0 + p_1x + \cdots + p_nx^n$ be an irreducible polynomial over a field F, so that $E = F[x]/\langle p(x)\rangle$ is a field. Let $\theta : F[x] \to E$ be the natural homomorphism and let $\alpha = \theta(x) = x + \langle p(x)\rangle$. Then:

(a) Any element of E can be represented, uniquely, in the form

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1},$$

where $a_0, a_1, \ldots, a_{n-1} \in F$.

With this representation of elements of E, addition is given by

$$\begin{aligned}&(a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}) + (b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1})\\ &\qquad = a_0 + b_0 + (a_1 + b_1)\alpha + \cdots + (a_{n-1} + b_{n-1})\alpha^{n-1},\end{aligned}$$

and multiplication is given by

$$\begin{aligned}&(a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1})(b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1})\\ &\qquad = (r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1}),\end{aligned}$$

where $r_0 + r_1x + \cdots + r_{n-1}x^{n-1} \in F[x]$ is the remainder in the division of $(a_0 + a_1x + \cdots + a_{n-1}x^{n-1})(b_0 + b_1x + \cdots + b_{n-1}x^{n-1})$ by $p(x)$.

(b) In E, we have

$$p(\alpha) = p_0 + p_1\alpha + \cdots + p_n\alpha^n = 0,$$

which means that α can be regarded as a root of the polynomial $p(x)$ with this representation of the elements of E.

Proof Since $p(x)$ is irreducible over $F[x]$, E is a field by Proposition 1.5.

(a) Firstly, $p(x)$ is irreducible, and so it has degree at least one. Thus, for $a \in F$ we have $\theta(a) = a + \langle p(x) \rangle$, so that

$$\theta(ax^m) = \theta(a)\theta(x)^m = a\alpha^m.$$

Hence if $f(x) = a_0 + a_1 x + \cdots + a_m x^m$ with $a_0, a_1, \ldots, a_m \in F$, then

$$\theta(f(x)) = a_0 + a_1\alpha + \cdots + a_m\alpha^m = f(\alpha).$$

Let $f(x)$ be a polynomial in $F[x]$. We use the Division Algorithm to write

$$f(x) = q(x)p(x) + r(x)$$

where $\deg r(x) < \deg p(x)$, or, equivalently,

$$f(x) = q(x)p(x) + r_0 + r_1 x + \cdots + r_{n-1}x^{n-1}$$

Some of the r_i could be 0.

for some $r_0, r_1, \ldots, r_{n-1} \in F$.

Proceeding as in Example 1.6 gives

$$f(x) - (r_0 + r_1 x + \cdots + r_{n-1}x^{n-1}) = q(x)p(x) \in \langle p(x) \rangle,$$

which can be written in the form

$$\begin{aligned} f(x) + \langle p(x) \rangle &= r_0 + r_1 x + \cdots + r_{n-1}x^{n-1} + \langle p(x) \rangle \\ &= r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1}. \end{aligned}$$

Thus every element of E can be represented uniquely in the form

$$r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1},$$

where $r_0, r_1, \ldots, r_{n-1} \in F$.

We can therefore regard E as a field with underlying set

$$\{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_0, a_1, \ldots, a_{n-1} \in F\}.$$

The homomorphism properties of θ ensure that the definition of addition and multiplication are correct when the elements of E are represented in this way.

(b) Since $\theta(p(x)) = 0$, we have
$p(\alpha) = p_0 + p_1\alpha + \cdots + p_n\alpha^n = \theta(p(x)) = 0.$ ■

The idea that the product of two elements is a remainder on division is one that you have already met when working in $\mathbb{Z}_n$. Just as we say in that case that multiplication is 'modulo n' and write $\pmod n$ in calculations to make this clear, we will describe multiplication in the quotient $F[x]/\langle p(x) \rangle$ as 'modulo $p(\alpha)$', and write $\pmod{p(\alpha)}$ where appropriate.

The next exercise illustrates Proposition 1.7.

Exercise 1.2

Let $p(x) = x^2 - x + 1$ be a polynomial in $\mathbb{Q}[x]$.

(a) Show that $p(x)$ is irreducible over $\mathbb{Q}[x]$.

(b) Let $E = \mathbb{Q}[x]/\langle p(x)\rangle$. Let $\alpha = x + \langle p(x)\rangle$ so that the elements of E can be written in the form $a + b\alpha$ where $a, b \in \mathbb{Q}$ as in Proposition 1.7. Write down the rule for multiplying two elements represented in this way.

In Example 1.6(a), we observed that the field $\mathbb{Q}[x]/\langle x^2 - 2\rangle$ is isomorphic to $\mathbb{Q}[\sqrt{2}]$. Similarly, $\mathbb{Q}[x]/\langle x^2 + 1\rangle \cong \mathbb{Q}[i]$; in fact, you saw this example just before Theorem 4.23 of Chapter 17. These isomorphisms enable us to carry out calculations using the more convenient notation $a + b\sqrt{2}$ and $a + bi$, respectively. The next worked exercise uses this approach when working over a finite field, as the mild abuse of notation improves readability.

Worked Exercise 1.8

(a) Show that $\mathbb{Z}_{13}[x]/\langle x^2 - 2\rangle$ and $\mathbb{Z}_{13}[x]/\langle x^2 - 5\rangle$ are fields.

(b) Show that the fields $\mathbb{Z}_{13}[x]/\langle x^2 - 2\rangle$ and $\mathbb{Z}_{13}[x]/\langle x^2 - 5\rangle$ are isomorphic.

Solution

(a) From the solution to Book C, Chapter 10, Exercise 1.2 we know that neither 2 nor 5 have square roots modulo 13, that is 2 and 5 are both quadratic non-residues of 13. Hence the two polynomials are irreducible and so $\mathbb{Z}_{13}[x]/\langle x^2 - 2\rangle$ and $\mathbb{Z}_{13}[x]/\langle x^2 - 5\rangle$ are fields by Proposition 1.5.

(b) By Proposition 1.7, the elements of $\mathbb{Z}_{13}[x]/\langle x^2 - 2\rangle$ can be written as $a + b\alpha$ where $a, b \in \mathbb{Z}_{13}$, with $\alpha^2 = 2$, and those of $\mathbb{Z}_{13}[x]/\langle x^2 - 5\rangle$ can be written as $c + d\beta$ with $\beta^2 = 5$ where $c, d \in \mathbb{Z}_{13}$.

To improve readability, we will mildly abuse notation by writing $\sqrt{2}$ for α and $\sqrt{5}$ for β.

We wish to define an isomorphism $\psi : \mathbb{Z}_{13}[x]/\langle x^2 - 2\rangle \to \mathbb{Z}_{13}[x]/\langle x^2 - 5\rangle$. In order to do this, we need to suitably define the image $\psi(a + b\sqrt{2})$ of an element of $\mathbb{Z}_{13}[x]/\langle x^2 - 2\rangle$.

Simply defining $\psi(a + b\sqrt{2}) = a + b\sqrt{5}$ will not work, since then

$$\psi(2) = \psi(\sqrt{2})\psi(\sqrt{2}) = \sqrt{5}\sqrt{5} = 5 \not\equiv 2 \pmod{13}.$$

So we need to define $\psi(\sqrt{2}) = r + s\sqrt{5}$ where $(r + s\sqrt{5})^2 \equiv 2 \pmod{13}$. In other words, we require r and s such that

$$(r + s\sqrt{5})^2 = r^2 + 5s^2 + 2rs\sqrt{5} \equiv 2 \pmod{13}.$$

This means that either r or s is zero. Since we know that 2 is a quadratic non-residue of 13, we have $r = 0$. Thus we require s such that $5s^2 \equiv 2 \pmod{13}$, that is so that $s^2 \equiv 3 \pmod{13}$. Since 3 is a quadratic residue of 13, and $4^2 \equiv 3 \pmod{13}$, we can choose $s = 4$.

Let $\psi(a + b\sqrt{2}) = a + 4b\sqrt{5}$ so that $\psi^{-1}(c + d\sqrt{5}) = c + 10d\sqrt{2}$ (as $4 \times 10 \equiv 1 \pmod{13}$). Therefore ψ is a bijection, and since $\psi(1) = 1$ we only have to show the isomorphism properties for addition and multiplication. For addition we have

$$\begin{aligned}\psi(a + b\sqrt{2}) + \psi(c + d\sqrt{2}) &= a + 4b\sqrt{5} + c + 4d\sqrt{5} \\ &= (a + c) + 4(b + d)\sqrt{5} \\ &= \psi(a + c + (b + d)\sqrt{2}) \\ &\equiv \psi((a + b\sqrt{2}) + (c + d\sqrt{2})) \pmod{13}.\end{aligned}$$

For multiplication we have

$$\begin{aligned}\psi(a + b\sqrt{2})\psi(c + d\sqrt{2}) &= (a + 4b\sqrt{5})(c + 4d\sqrt{5}) \\ &\equiv ac + 2bd + 4(ad + bc)\sqrt{5} \pmod{13}\end{aligned}$$

since $80 \equiv 2 \pmod{13}$, and

$$\begin{aligned}\psi((a + b\sqrt{2})(c + d\sqrt{2})) &= \psi(ac + 2bd + (ad + bc)\sqrt{2}) \\ &\equiv ac + 2bd + 4(ad + bc)\sqrt{5} \pmod{13}.\end{aligned}$$

Hence ψ is a field homomorphism.

Exercise 1.3

(a) Show that $\mathbb{Z}_7[x]/\langle x^2 - 5\rangle$ and $\mathbb{Z}_7[x]/\langle x^2 - 6\rangle$ are fields.

(b) Show that the fields $\mathbb{Z}_7[x]/\langle x^2 - 5\rangle$ and $\mathbb{Z}_7[x]/\langle x^2 - 6\rangle$ are isomorphic.

This subsection concludes with an exercise targeting the fields of rational and complex numbers.

Exercise 1.4

(a) Show that $\mathbb{Q}[x]/\langle x - 2\rangle \cong \mathbb{Q}$.

(b) Let $p(x)$ be an irreducible polynomial over $\mathbb{C}[x]$. Show that $\mathbb{C}[x]/\langle p(x)\rangle \cong \mathbb{C}$.

1.2 Characteristic of a field

The fields you have met so far either contain $\mathbb{Q}$ or $\mathbb{Z}_p$, where p is a prime, or they are obtained from $\mathbb{Q}$ or $\mathbb{Z}_p$ by a construction such as in Proposition 1.5. We will now show that this is always the case. We start by defining the *characteristic* of a field.

Definition 1.9 *Characteristic of a field*

If the order of 1 in the additive group $(F, +)$ of a field F is infinite, then F is said to have **characteristic zero**. If the order of 1 in the group $(F, +)$ is finite and equal to m, we say that the field F has **characteristic** m.

Next we prove that the possibilities for the non-zero characteristic of a field are rather restricted.

Theorem 1.10

The characteristic of any field is either prime or zero.

Proof Assume that F is a field of some positive characteristic m. That is, in F we have $0 = 1 + 1 + \cdots + 1$ where the sum contains exactly m terms, and m is the smallest positive integer for which this equality holds.

Suppose that m is not a prime and let $m = m_1 m_2$ where m_1 and m_2 are positive integers smaller than m. Then, in F we would have

$$0 = (\underbrace{1 + 1 + \cdots + 1}_{m_1 \text{ times}})(\underbrace{1 + 1 + \cdots + 1}_{m_2 \text{ times}}).$$

However, this gives two non-zero elements of F whose product is zero, which contradicts the fact that a field cannot contain zero divisors.

Therefore if the characteristic of F is finite, it must be prime. Summing up, the characteristic of any field is zero or prime. ■

A key fact that you should be aware of is that in a field of characteristic p, any multiple of $p \cdot 1$, that is, any multiple of $1 + \cdots + 1$ where the sum has p terms, is zero.

The next exercise asks you to identify a particular subfield of a field of prime characteristic.

Exercise 1.5

Let F be a field of prime characteristic p. Consider the subset

$$K = \{0,\ 1,\ 1+1,\ \ldots,\ \underbrace{1+1+\cdots+1}_{p-1 \text{ times}}\} \subset F$$

that contains all possible sums of the multiplicative identity 1. Show that K is a subfield of F.

Motivated by Exercise 1.5, we now state and prove a theorem about the smallest subfields of fields of prime and zero characteristic.

Theorem 1.11

If a field F has prime characteristic p, then F contains a subfield isomorphic to $\mathbb{Z}_p$, and this is the smallest subfield of F with respect to inclusion, that is, it is contained in every other subfield of F.

Similarly, if F has characteristic zero, then F contains a subfield isomorphic to $\mathbb{Q}$, and this is the smallest subfield of F with respect to inclusion.

Proof Let F have prime characteristic p. Consider the elements 0, 1, $1+1$, $\ldots$, up to $1+1+\cdots+1$ where the sum has $p-1$ terms. We know from the solution to Exercise 1.5 that the set formed by these p elements forms a subfield K of F and that the mapping $\theta : \mathbb{Z}_p \to K$ that maps the residue class $0 \pmod p$ onto the element $0 \in K$ and, for $1 \leq k \leq p-1$, the residue class $k \pmod p$ onto the sum $1+1+\cdots+1 \in K$ with k terms is a homomorphism. Hence, by the solution to Exercise 1.1, θ is an isomorphism of fields since $K = \mathrm{Im}(\theta)$.

To prove minimality, observe that any subfield of F must contain the multiplicative identity 1. But since, as we have seen above, this multiplicative unit generates the subfield K that is isomorphic to $\mathbb{Z}_p$, we conclude that K is the smallest subfield of F, that is, it is contained in any other subfield of F.

Let F have characteristic 0 and consider the elements $1+1$, $1+1+1$, and so on, together with their additive inverses. By an argument similar to that above, F must contain an isomorphic copy of the integers. Since F is a field, it must also contain multiplicative inverses to all the non-zero elements that are in this isomorphic copy. That is, F must contain an isomorphic copy of all integers m and all reciprocals $\frac{1}{n}$ of non-zero elements n. But again, since F is a field, it must contain the product of any two of its elements, and hence it must contain all fractions $m \cdot \frac{1}{n} = \frac{m}{n}$. We conclude that F contains an isomorphic copy of $\mathbb{Q}$. ■

The smallest subfield of a field F of prime or zero characteristic, referred to in Theorem 1.11, is often called the **prime field** of F.

Theorem 1.11 also shows that, up to isomorphism, there is only one field of p elements for any given prime p. Until now we have used the notation $\mathbb{Z}_p$ for this field and we will continue to do so. There is, however, an alternative notation, namely, GF_p, read as the **Galois field of p elements**. This notation honours Evariste Galois, who was a major influence on the development of the theory of finite fields.

The next result is often useful when working in a field of characteristic p. It is sometimes known as the **Idiot's Binomial Theorem**.

Theorem 1.12

Let F be a field of characteristic p. Then

$$(u+v)^p = u^p + v^p$$

for all $u, v \in F$.

Remember that $\binom{p}{j} = \dfrac{p!}{(p-j)!j!}$.

Proof Using the binomial theorem to expand $(u+v)^p$ yields

$$(u+v)^p = u^p + \binom{p}{1}u^{p-1}v + \cdots + \binom{p}{j}u^{p-j}v^j + \cdots + \binom{p}{p-1}uv^{p-1} + v^p \ .$$

However, here every binomial coefficient $\binom{p}{j}$ is to be interpreted as the sum $1+1+\cdots+1$ with $\binom{p}{j}$ terms. Now, for each j such that $1 \leq j \leq p-1$, we have $p \mid p!$, $p \nmid (p-j)!$ and $p \nmid j!$, and so $\binom{p}{j}$ is a multiple of p. Thus $\binom{p}{j} = 0$ in the field F. Hence in the field F the above binomial expansion reduces to $(u+v)^p = u^p + v^p$. ■

A word of caution regarding prime characteristic: you should not think that a field of prime characteristic is necessarily finite. For an example of an infinite field of finite characteristic, you are invited to consider the field of fractions of the ring $\mathbb{Z}_p[x]$ for an arbitrary prime p.

This section ends with two results relating field isomorphisms and characteristic. The first result proves that isomorphic fields have the same characteristic.

Proposition 1.13

Let F_1 and F_2 be fields and suppose that there is a non-zero field homomorphism $\phi : F_1 \to F_2$. Then F_1 and F_2 have the same characteristic.

Proof Suppose that F_1 has characteristic 0. Then, by Theorem 1.11, F_1 has a subfield isomorphic to $\mathbb{Q}$. Since, by Proposition 1.4, F_2 has a subfield isomorphic to F_1, it also has a subfield isomorphic to $\mathbb{Q}$. Thus 1 has infinite order in F_2 and F_2 has characteristic 0.

Suppose that F_1 has finite characteristic p for some prime p. Theorem 1.11 then tells us that F_1 has a subfield isomorphic to $\mathbb{Z}_p$. As above, since F_2

has a subfield isomorphic to F_1, it also has a subfield isomorphic to $\mathbb{Z}_p$. Thus 1 has finite order p in F_2 and F_2 has characteristic p. ■

The final result of this section tells us that any isomorphism between subfields of $\mathbb{C}$ is the identity map on $\mathbb{Q}$.

Theorem 1.14

Let F_1 and F_2 be subfields of $\mathbb{C}$ and let $\phi : F_1 \to F_2$ be an embedding. Then $\phi(x) = x$ for every $x \in \mathbb{Q}$.

Proof By Theorem 1.11, $\mathbb{Q}$ is a subfield of both F_1 and F_2. Since ϕ must preserve the additive and the multiplicative identities, we have $\phi(0) = 0$ and $\phi(1) = 1$. By the definition of a field homomorphism, for any positive integer m we have

$$\phi(m) = \phi(\underbrace{1 + \cdots + 1}_{m \text{ times}}) = \underbrace{\phi(1) + \cdots + \phi(1)}_{m \text{ times}} = m.$$

This means that $\phi(m) = m$ for every positive integer m. As ϕ preserves additive inverses, $\phi(-m) = -m$ and hence $\phi(n) = n$ for all integers n. Since ϕ also preserves multiplicative inverses, for each non-zero integer n we have

$$\phi\left(\frac{1}{n}\right) = \frac{1}{\phi(n)} = \frac{1}{n}.$$

Finally, from the properties of a field homomorphism we deduce that for all integers m, n with $n \neq 0$ we have

$$\phi\left(\frac{m}{n}\right) = \phi\left(m \cdot \frac{1}{n}\right) = \phi(m) \cdot \phi\left(\frac{1}{n}\right) = \frac{m}{n}.$$

This shows that $\phi(x) = x$ for all $x \in \mathbb{Q}$. ■

2 Field extensions

This section introduces field extensions. We start with the definition of a field extension and then look at some examples. We will see how regarding an extension of a field F as a vector space over F allows us to determine some key properties of extensions – in particular their *degree*, which is, in a sense that will be made clear, a measure of dimension. We then go on to study some particular types of extension.

Definition 2.1 *Extension of a field*

A field E is an **extension of a field** F if F is a subfield of E.

In the next two examples we will use the word 'identify'. We used it before when regarding integers as a subset of $\mathbb{Q}$ in Chapter 17, and also when regarding a ring R as a subring of the ring of polynomials $R[x]$ by 'identifying' the constant polynomials with R. When we say that we *identify* two isomorphic objects, such as two fields, we mean that we regard them as not just isomorphic, but equal. The main place where we use this idea is in connection with the natural homomorphism from a polynomial ring over a field F to a quotient ring over a maximal ideal M; in this setting, we identify F with its image $\{k + M : k \in F\}$ under the natural homomorphism.

Example 2.2 *Some field extensions*

(a) $\mathbb{C}$ is an extension of $\mathbb{Q}$.

(b) Let $E = \{a + b\alpha : a, b \in \mathbb{Q}\}$, where $\alpha^2 = 2$. Define addition and multiplication on E by

$$(a + b\alpha) + (c + d\alpha) = (a + c) + (b + d)\alpha,$$
$$(a + b\alpha)(c + d\alpha) = ac + 2bd + (ad + bc)\alpha.$$

Then, as you saw in Example 1.6(a), E is a field isomorphic to $\mathbb{Q}[x]/\langle x^2 - 2\rangle$. The subset $F = \{a + b\alpha : a \in \mathbb{Q}, b = 0\}$ of E is isomorphic to $\mathbb{Q}$ and E is an extension of F. In this type of example we usually identify $\mathbb{Q}$ with F (as we do when regarding the integers as a subset of $\mathbb{Q}$) and say that E is an extension of $\mathbb{Q}$.

(c) Let $E = \{a + b\alpha : a, b \in \mathbb{Z}_7\}$, where $\alpha^2 = 3$. Define addition and multiplication on E by

$$(a + b\alpha) + (c + d\alpha) = (a + c) + (b + d)\alpha \pmod 7,$$
$$(a + b\alpha)(c + d\alpha) = ac + 3bd + (ad + bc)\alpha \pmod 7.$$

Again, by Example 1.6(b), E is a field isomorphic to $\mathbb{Z}_7[x]/\langle x^2 - 3\rangle$. The subset $F = \{a + b\alpha : a \in \mathbb{Z}_7, b = 0\}$ of E is isomorphic to $\mathbb{Z}_7$ and E is an extension of F. As in the previous example, we identify $\mathbb{Z}_7$ with F and say that E is an extension of $\mathbb{Z}_7$.

2.1 Vector spaces

One way to look at field extensions is to think in terms of vector spaces. In order to do this we need to consider vector spaces where the scalars come from a general field. In your previous studies you have met vector spaces where the scalars are real numbers. However, the definition of a vector space over $\mathbb{R}$ and the key notions only make use of the field properties of $\mathbb{R}$, and so they are easily extended to the case where the scalars come from an arbitrary field.

Definition 2.3 *Vector spaces over general fields*

A **vector space over a field** F consists of a set V of elements and two operations – vector addition and scalar multiplication – such that the following axioms hold.

Axioms for vector addition:

A1 Closure of vector addition For all $\mathbf{v}_1, \mathbf{v}_2 \in V$,

$$\mathbf{v}_1 + \mathbf{v}_2 \in V.$$

A2 Additive identity There is a vector $\mathbf{0} \in V$ such that for each $\mathbf{v} \in V$,

$$\mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v} = \mathbf{v}.$$

A3 Inverses For each $\mathbf{v} \in V$, there is an element $-\mathbf{v}$ (its additive inverse) such that

$$\mathbf{v} + (-\mathbf{v}) = (-\mathbf{v}) + \mathbf{v} = \mathbf{0}.$$

A4 Associativity of vector addition For all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V$,

$$(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3).$$

A5 Commutativity For all $\mathbf{v}_1, \mathbf{v}_2 \in V$,

$$\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1.$$

Axioms for scalar multiplication:

S1 Closure under scalar multiplication For all $\mathbf{v} \in V$ and $\alpha \in F$,

$$\alpha\mathbf{v} \in V.$$

S2 Associativity of scalar multiplication For all $\mathbf{v} \in V$ and $\alpha, \beta \in F$,

$$\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}.$$

S3 Scalar identity For all $\mathbf{v} \in V$,

$$1\mathbf{v} = \mathbf{v}.$$

Axioms combining addition and multiplication:

D1 Distributive law For all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\alpha \in F$,

$$\alpha(\mathbf{v}_1 + \mathbf{v}_2) = \alpha\mathbf{v}_1 + \alpha\mathbf{v}_2.$$

D2 Distributive law For all $\mathbf{v} \in V$ and $\alpha, \beta \in F$,

$$(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}.$$

Observe that axioms A1–A5 say that $(V, +)$ is an abelian group.

With Definition 2.3 in mind, we revisit Example 2.2 to see how we can view an extension of a field as a vector space over that field.

Example 2.4

(a) Consider the field of complex numbers $\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}$. We can regard $\mathbb{C}$ as a vector space over $\mathbb{R}$. We can also regard it as a vector space over $\mathbb{Q}$ where scalar multiplication is just multiplication of a complex number by a rational number.

(b) Let us revisit the field $E = \{a + b\alpha : a, b \in \mathbb{Q}\}$ in Example 2.2(b), where $\alpha^2 = 2$ and addition and multiplication on E are given by

$$(a + b\alpha) + (c + d\alpha) = (a + c) + (b + d)\alpha,$$
$$(a + b\alpha)(c + d\alpha) = ac + 2bd + (ad + bc)\alpha.$$

We can also view E as a vector space over $\mathbb{Q}$. Since E is a field, $(E, +)$ is an abelian group, and so axioms A1–A5 are satisfied. Scalar multiplication is defined using the definition of multiplication above with $b = 0$, so that for $a \in \mathbb{Q}$ and $c + d\alpha \in E$,

$$a(c + d\alpha) = ac + ad\alpha.$$

Then axioms S1–S3 and D1–D2 are satisfied because E is a field.

(c) Let us look at the field $E = \{a + b\alpha : a, b \in \mathbb{Z}_7\}$ in Example 2.2(c), where $\alpha^2 = 3$ and addition and multiplication on E are given by

$$(a + b\alpha) + (c + d\alpha) = (a + c) + (b + d)\alpha \pmod 7,$$
$$(a + b\alpha)(c + d\alpha) = ac + 3bd + (ad + bc)\alpha \pmod 7.$$

We can also view E as a vector space over $\mathbb{Z}_7$. Since E is a field, axioms A1–A5 are satisfied. Scalar multiplication is defined using the definition of multiplication above with $b = 0$, so that, again since E is a field, axioms S1–S3 and D1–D2 are satisfied.

The previous examples generalise in a way you might have expected.

Proposition 2.5 *Field extensions as vector spaces*

If E is an extension of a field F, then E is a vector space over F.

Proof Since E is a field, $(E, +)$ is an abelian group, so the (abelian group) axioms A1–A5 are satisfied. Scalar multiplication is defined using the definition of multiplication in F, that is, for $m \in F$ and $v \in E$, the multiple of the vector v by the scalar m is defined to be the product mv in E. Then, since E is a field, axioms S1–S3 and D1–D2 are satisfied. ■

You might object that when we think of a field extension as a vector space we 'hide' some information. But one of the skills needed in mathematics is the ability to view a mathematical object in different ways, and by ignoring some details it is possible to understand other details more clearly.

In our current context we will use the vector space 'view' to introduce the idea of *degree* of a field extension. We first need to recall the relevant notions in a vector space.

Definition 2.6 *Linear independence and basis*

Let V be a vector space over a field F.

(a) Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in V$. A **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is a vector of the form

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$.

(b) Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a finite set of vectors in a vector space V. Then the **span** $\langle S \rangle$ of S is the set of all possible linear combinations

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$; that is,

$$\langle S \rangle = \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k : \alpha_1, \alpha_2, \ldots, \alpha_k \in F\}.$$

We say that the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ **spans** $\langle S \rangle$ or is a **spanning set** for $\langle S \rangle$, and that $\langle S \rangle$ is the set **spanned** by S.

(c) A finite set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in a vector space V is a **linearly dependent** set if there exist $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$, *not all zero*, such that

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}.$$

(d) A finite set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is **linearly independent** if it is not linearly dependent; that is, if

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$$

only when $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.

(e) A *basis* for a vector space V is a linearly independent set of vectors that is a spanning set for V.

(f) Let V be a vector space. Then V has **finite dimension** if it contains a finite set of vectors S that forms a basis for V. If no such set exists, then V has **infinite dimension**.

You are now invited to test your understanding of Definition 2.6 on a field you have encountered before.

Exercise 2.1

Consider again the field $E = \{a + b\alpha : a, b \in \mathbb{Q}\}$ in Example 2.2(b), with $\alpha^2 = 2$ and with addition and multiplication defined by

$$(a + b\alpha) + (c + d\alpha) = (a + c) + (b + d)\alpha,$$
$$(a + b\alpha)(c + d\alpha) = ac + 2bd + (ad + bc)\alpha,$$

viewed as an extension of the field $F = \{a + b\alpha : a, b \in \mathbb{Q},\ b = 0\} \cong \mathbb{Q}$. Write down a basis for E, regarded as a vector space over F.

Not all vector spaces have finite dimension. For example, $\mathbb{C}$ as a vector space over $\mathbb{Q}$ does not have a finite basis, and so it has infinite dimension. When a vector space does have a finite basis, the following Basis Theorem (which you may have encountered in your previous studies for vector spaces over $\mathbb{R}$, and the proof of which we omit) tells us that every basis has the same number of elements.

Theorem 2.7 *Basis Theorem*

Let V be a finite-dimensional vector space. Then every basis for V contains the same number of vectors.

This enables us to define the degree of a field extension.

Definition 2.8 *Degree of a field extension*

Let E be an extension of a field F. If E has a finite basis as a vector space over F, we say that the **degree** of the extension is its dimension as a vector space over F, that is, the number of vectors in any basis for the space. We denote the degree by $[E : F]$. If E does not have a finite basis as a vector space over F, then we say that the extension is infinite and we write $[E : F] = \infty$.

Viewing field extensions as vector spaces enables us to show in an elegant way that the order of a finite field is a prime power.

Theorem 2.9

The order of a finite field is a power of a prime.

Proof Let F be a finite field. By the previous results we know that F has characteristic p for some prime p, and it contains (an isomorphic copy of) $\mathbb{Z}_p$ as a subfield.

Now, by Proposition 2.5, the field F is a vector space over the field $\mathbb{Z}_p$. Since F is finite, it has, as a vector space over $\mathbb{Z}_p$, a finite dimension, say, n. Hence, there exists a basis B for the vector space F over $\mathbb{Z}_p$ such that $|B| = n$, say $B = \{\beta_1, \beta_2, \ldots, \beta_n\} \subset F$.

Since B is a basis for F, for every element $\alpha \in F$ there exists a *unique* n-tuple $(c_1, c_2, \ldots, c_n)$ of elements of $\mathbb{Z}_p$ such that

$$\alpha = c_1\beta_1 + c_2\beta_2 + \cdots + c_n\beta_n.$$

Conversely, every linear combination of elements of B with coefficients from $\mathbb{Z}_p$ is an element of F.

This means that the mapping $\alpha \mapsto (c_1, c_2, \ldots, c_n)$ is a bijection from F into the set $(\mathbb{Z}_p)^n$ of all n-tuples of elements of $\mathbb{Z}_p$.

Therefore, $|F| = |(\mathbb{Z}_p)^n| = p^n$. In other words, since this bijection enables us to identify each element of F with an n-tuple of elements of $\mathbb{Z}_p$, and since there are p^n distinct n-tuples of elements of $\mathbb{Z}_p$, we conclude that F must have p^n elements. ■

2.2 Extensions of fields by elements

If E is an extension of a field F and if α is an element of E that may or may not be in F, it makes sense to study the subfields of E that contain F and the element α. Such a point of view leads to a rich theory.

Definition 2.10 $F(\alpha)$

Let E be an extension of a field F and let $\alpha \in E$. We denote by $F(\alpha)$ the smallest subfield of E (with respect to inclusion) that contains both F and α.

A consequence of Definition 2.10 is that $F(\alpha)$ contains, at the very least, the set of all polynomials $a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_r\alpha^r$ for all $r \geq 0$, where $a_i \in F$ for all i such that $0 \leq i \leq r$.

Note the use of round brackets in the notation $F(\alpha)$. Until now we have been using square brackets to denote an extension of a ring, for example $\mathbb{Z}[\sqrt{2}]$ or $\mathbb{Q}[\sqrt{2}]$. In the latter case, the ring is in fact a field (see Example 2.2(a)). In cases like this we will use the round bracket notation from now on.

An extension $F(\alpha)$ is of one of two kinds, determined by whether α is a root of a polynomial in $F[x]$.

Definition 2.11 *Algebraic and transcendental element*

Let the field E be an extension of a field F. An element $\alpha \in E$ is said to be **algebraic over** F if there is a non-zero polynomial $h(x)$ in $F[x]$ with $h(\alpha) = 0$. An element $\alpha \in E$ is called **transcendental over** F if there is no polynomial $h(x)$ in $F[x]$ with $h(\alpha) = 0$.

We also want to consider extensions by several elements. We might, for example, want to extend $\mathbb{Q}$ by both $\sqrt{2}$ and $\sqrt{3}$ if we wanted all the roots of the polynomial $(x^2 - 2)(x^2 - 3)$ to be in the extension. Such a case is covered in the following definition.

Definition 2.12 $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$

Let E be an extension of a field F and let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subseteq E$. We denote by $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ the smallest subfield of E (with respect to inclusion) that contains both F and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$.

The field $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ can be obtained by iterating the one-element extensions defined in Definition 2.10. The next proposition shows how to do this for $n = 2$ only; you are invited to prove the corresponding statement for an arbitrary positive integer n by induction.

Proposition 2.13

Let E be an extension of a field F and let $\alpha_1, \alpha_2 \in E$. Then $F(\alpha_1, \alpha_2) = (F(\alpha_1))(\alpha_2)$.

Proof The statement follows from the fact that $F(\alpha_1, \alpha_2)$ is the smallest subfield of E containing F and $\{\alpha_1, \alpha_2\}$, and $(F(\alpha_1))(\alpha_2)$ is the smallest subfield of E containing $F(\alpha_1)$ and α_2. ■

The next definition introduces the appropriate terminology for different kinds of extensions of the form $F(\alpha)$.

Definition 2.14 *Algebraic and transcendental extension*

An extension $F(\alpha)$ of a field F is called an **algebraic extension** if α is an algebraic element over F. Analogously, $F(\alpha)$ is called a **transcendental extension** if α is transcendental over F.

The next proposition enables us to generalise the definition of algebraic extension.

Proposition 2.15

Let E be an extension of a field F and let $\alpha \in E$ be algebraic over F. Let $\beta \in F(\alpha)$. Then β is algebraic over F.

Proof Since α is algebraic over F, there is a polynomial $p(\alpha) \in F[x]$ of degree n, say, such that $p(\alpha) = 0$. The equation $p(\alpha) = 0$ expresses a linear dependence among the elements $1, \alpha, \alpha^2, \ldots, \alpha^n$, which implies that $S = \{1, \alpha, \alpha^2, \ldots, \alpha^n\}$ is a linearly dependent set in the vector space $F(\alpha)$ over F. By Definition 2.10, every element of $F(\alpha)$ is of the form $a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_r\alpha^r$ for some $r \geq 0$, where $a_i \in F$ for all i, $0 \leq i \leq r$. We conclude that S is a spanning set for $F(\alpha)$ as a vector space over F.

It follows that there is a subset B of S that is a basis for $F(\alpha)$ over F. In particular, $F(\alpha)$ is a finite-dimensional extension of F. Suppose that $[F(\alpha) : F] = m$ and consider an element $\beta \in F(\alpha)$. Then by Theorem 2.7 $\{1, \beta, \ldots, \beta^m\}$ is a linearly dependent set, so there exist $b_i \in F$, for $0 \leq i \leq m$, such that $b_0 + b_1\beta + \cdots + b_m\beta^m = 0$ and not all the b_i are zero. Hence β is algebraic over F, as required. ■

Proposition 2.15 prompts the following definition for arbitrary extensions of fields.

Definition 2.16 *Algebraic and transcendental extension (general)*

An extension E of a field F is called an **algebraic extension** if every element in E is algebraic over F. Otherwise (that is, if E contains a transcendental element), the extension is called **transcendental**.

It is important to observe that Proposition 2.15 means that this definition is consistent with Definition 2.14.

You have already encountered both algebraic and transcendental extensions. The next example revisits some of them.

Example 2.17 *Algebraic and transcendental field extensions*

The fields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(i)$ are both algebraic extensions of $\mathbb{Q}$. The field $\mathbb{Z}_7[x]/\langle x^2-3\rangle$ in Example 1.6(b) is an algebraic extensions of $\mathbb{Z}_7$, the field $\mathbb{Q}[x]/\langle x^2-x-1\rangle$ in Exercise 1.2 is an algebraic extension of $\mathbb{Q}$, and the fields $\mathbb{Z}_{13}[x]/\langle x^2-2\rangle$ and $\mathbb{Z}_{13}[x]/\langle x^2-5\rangle$ in Worked Exercise 1.8 are algebraic extensions of $\mathbb{Z}_{13}$.

Example 2.18 *The transcendental extensions $\mathbb{Q}(e)$ and $\mathbb{Q}(\pi)$*

As a consequence of some results in advanced mathematical analysis, the Euler number e and the number π are real numbers that are not roots of any polynomial with rational coefficients. In our terminology, e and π are transcendental over $\mathbb{Q}$, so the extensions $\mathbb{Q}(e)$ and $\mathbb{Q}(\pi)$ are transcendental.

Charles Hermite proved that e is transcendental in 1873, and in 1882 Ferdinand von Lindemann proved that π is transcendental by showing that e^q is transcendental for every non-zero algebraic number q. Since then, alternative proofs of the transcendence of these two numbers have been found.

2.3 Algebraic extensions

This subsection investigates algebraic extensions in more detail. First we show that we can construct an algebraic extension of a field F from a given irreducible polynomial over F. We use Proposition 1.7 and construct the extension from the cosets in $F[x]/\langle p(x)\rangle$, and we identify the elements of F with the cosets $k+\langle p(x)\rangle$ for $k\in F$.

Consider an irreducible polynomial $p(x)$ of degree n over a field F. In Proposition 1.7 you saw that the elements of the field $E=F[x]/\langle p(x)\rangle$ may be represented as $a_0+a_1\alpha+\cdots+a_{n-1}\alpha^{n-1}$, where $\alpha=x+\langle p(x)\rangle$ and $a_0,a_1,\ldots,a_{n-1}\in F$. Recall that, in this representation, addition in E is given by

$$\begin{aligned}(a_0+a_1\alpha+\cdots+a_{n-1}\alpha^{n-1})+(b_0+b_1\alpha+\cdots+b_{n-1}\alpha^{n-1})\\ =a_0+b_0+(a_1+b_1)\alpha+\cdots+(a_{n-1}+b_{n-1})\alpha^{n-1},\end{aligned}$$

and multiplication is given by

$$(a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1})(b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}) \\ = (r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1}),$$

where $r(x) = r_0 + r_1x + \cdots + r_{n-1}x^{n-1}$ is the remainder when the product $(a_0 + a_1x + \cdots + a_{n-1}x^{n-1})(b_0 + b_1x + \cdots + b_{n-1}x^{n-1})$ is divided by $p(x)$ in $F[x]$.

Regarding E as a vector space over F gives us substantial information about the extension E of F.

Theorem 2.19

Let F be a field and let $p(x)$ be an irreducible polynomial of degree n in $F[x]$. Let $E = F[x]/\langle p(x)\rangle$ and let $\alpha = x + \langle p(x)\rangle$. Let us represent E in the form $E = \{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_0, a_1, \ldots, a_{n-1} \in F\}$. Then:

(a) E is an extension of F

(b) $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ is a basis for E as a vector space over F

(c) the degree $[E : F]$ of the field extension E over F is n

(d) $p(\alpha) = 0$

(e) $E = F(\alpha)$ and it is an algebraic extension over F.

Proof

(a) The subfield $\{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_0 \in F, a_1 = \cdots = a_{n-1} = 0\}$ of E may be identified with F, which shows that E is an extension of F.

(b) Consider the set $B = \{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ and regard E as a vector space over F. Because of our chosen representation of E, B is a spanning set.

Suppose that

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} = 0 \in E\,.$$

This is equivalent to writing

$$a_0 + a_1(x + \langle p(x)\rangle) + \cdots + a_{n-1}(x + \langle p(x)\rangle)^{n-1} = 0 \in E\,,$$

which is the same as

$$a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + \langle p(x)\rangle = 0 \in E\,,$$

or, equivalently,

$$a_0 + a_1x + \cdots + a_{n-1}x^{n-1} \in \langle p(x)\rangle\,.$$

But $\deg p(x) = n$, so the only polynomial of degree lower than n in $\langle p(x)\rangle$ is the zero polynomial. Therefore

$$a_0 + a_1x + \cdots + a_{n-1}x^{n-1} = 0 \in F[x]\,.$$

But this means that $a_0 = a_1 = \cdots = a_{n-1} = 0$.

In the language of vector spaces, we have shown that B is also a linearly independent set. Thus B is a basis for E as a vector space over F.

(c) By part (b), B is a basis for E as a vector space over F, and so $[E : F] = n$.

(d) This follows from Proposition 1.7.

(e) This follows from the fact that every element of E can be written in the form $a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}$ for suitable $a_0, a_1, \ldots, a_{n-1} \in F$, and from $p(\alpha) = 0$. ■

The algebraic extension $F(\alpha)$ of F described in Theorem 2.19 is sometimes said to be obtained by **adjoining** α to F.

We can use Theorem 2.19 to show that for any non-constant polynomial $f(x)$, not necessarily irreducible, over a field F, we can construct an extension of F that contains a root of $f(x)$.

Corollary 2.20

Let F be a field and let $f(x)$ be a non-constant polynomial in $F[x]$. Then there is an extension E of F that contains an element α such that $f(\alpha) = 0$.

Proof Since $f(x)$ is not a constant, there is an irreducible polynomial $h(x) \in F[x]$ such that $f(x) = h(x)g(x)$ for some $g(x) \in F[x]$. By Theorem 2.19, there are an extension E of F and an element $\alpha \in E$ such that $h(\alpha) = 0$. Since $f(x) = h(x)g(x)$, $f(\alpha) = 0$, and so E is the required extension of F. ■

We will now reverse our line of thought and show that one can start with an algebraic element α in an extension E of a field F and find an irreducible polynomial that gives rise to the extension $F(\alpha)$.

Theorem 2.21

Let E be an extension of a field F and suppose that $\alpha \in E \backslash F$ is algebraic over F. Then:

(a) there is an irreducible polynomial $h(x) \in F[x]$ such that $h(\alpha) = 0$ and $F[x]/\langle h(x)\rangle \cong F(\alpha)$

(b) there is a unique irreducible monic polynomial $m(x) \in F[x]$ such that $m(\alpha) = 0$ and $F[x]/\langle m(x)\rangle \cong F(\alpha)$

(c) $[F(\alpha) : F] = \deg m(x) = \deg h(x)$.

Proof

(a) Since α is algebraic over F, there is a polynomial $h(x)$ such that $h(\alpha) = 0$. If $h(x)$ is reducible, it can be written as a product $h_1(x)h_2(x)$ where both $h_1(x)$ and $h_2(x)$ have lower degree than $h(x)$. Since $h(\alpha) = 0$, either $h_1(\alpha) = 0$ or $h_2(\alpha) = 0$. Since this process must stop, we can assume, without loss of generality, that $h(x)$ is irreducible.

Let $f_\alpha : F[x] \to F(\alpha)$ be the evaluation mapping, which is defined by $f_\alpha(r(x)) = r(\alpha)$ for every polynomial $r(x) \in F[x]$. This map can be shown to be a homomorphism by an argument similar to the solution to Exercise 4.7 of Chapter 17 and the discussion following it. This homomorphism is onto, since every element of $F(\alpha)$ can be expressed as a finite sum of elements of the form $c\alpha^k$ and so can be written in the form $f_\alpha(s(x))$ for some polynomial $s(x) \in F[x]$.

To complete the proof, we need to show that $\mathrm{Ker}(f_\alpha) = \langle h(x)\rangle$. It is clear that $\langle h(x)\rangle \subseteq \mathrm{Ker}(f_\alpha)$, since $h(\alpha) = 0$ and so any polynomial in $\langle h(x)\rangle$ must map to 0. Since $h(x)$ is irreducible, we know from Theorem 4.22 of Chapter 17 that $\langle h(x)\rangle$ is a maximal ideal. Hence $\langle h(x)\rangle = \mathrm{Ker}(f_\alpha)$ and so by the First Isomorphism Theorem for rings we have $F[x]/\langle h(x)\rangle \cong F(\alpha)$.

(b) Let $m(x) = ah(x)$ where a is the inverse of the leading coefficient of $h(x)$. Then $m(x)$ is monic and irreducible and $\langle m(x)\rangle = \langle h(x)\rangle$. It remains to show that $m(x)$ is unique.

Suppose that $g(x)$ is a monic irreducible polynomial with $g(\alpha) = 0$ and such that $F[x]/\langle g(x)\rangle \cong F(\alpha)$. Then it follows from Theorem 2.19 that

$$\deg g(x) = [F(\alpha) : F] = \deg m(x)\,.$$

Since both $g(x)$ and $m(x)$ are monic polynomials, the polynomial $k(x) = g(x) - m(x)$ has lower degree and satisfies $k(\alpha) = 0$. This is only possible if $k(x)$ is the zero polynomial, that is, if $g(x) = m(x)$ as required.

(c) This follows directly from Theorem 2.19. ■

The result of Theorem 2.21(b) suggests making the following definition.

Definition 2.22 *Minimal polynomial*

Let E be an extension of a field F and suppose that $\alpha \in E\backslash F$ is algebraic over F. A non-zero polynomial $h(x) \in F[x]$ is called the **minimal polynomial** for α if it is monic and a polynomial of lowest degree in $F[x]$ such that $h(\alpha) = 0$.

By Theorem 2.21(b), for each $\alpha \in E$ there is a unique minimal polynomial for α, so we refer to *the* minimal polynomial of an algebraic element over a field.

The next exercise clarifies the relation between irreducible monic polynomials and minimal polynomials.

Exercise 2.2

Let E be an extension of a field F and suppose that $\alpha \in E \backslash F$ is algebraic over F. Let $h(x)$ be a monic polynomial that is irreducible over F and such that $h(\alpha) = 0$. Show that $h(x)$ is the minimal polynomial for α.

The solution to Exercise 2.2 enables us to prove the following corollary to Theorem 2.21, where we define a field extension starting from an irreducible polynomial rather than from an algebraic element.

Corollary 2.23

Let E be an extension of a field F and suppose that $h(x)$ is an irreducible polynomial in $F[x]$ and that $\alpha \in E$ with $h(\alpha) = 0$. Then $F(\alpha) \cong F[x]/\langle h(x) \rangle$.

Proof Let $m(x)$ be the polynomial obtained by dividing $h(x)$ through by the leading coefficient. Then $m(x)$ is monic and irreducible and $m(\alpha) = 0$. Then, by the solution to Exercise 2.2, $m(x)$ is the minimal polynomial for α. Since $\langle m(x) \rangle = \langle h(x) \rangle$ the result follows by Theorem 2.21. ■

The next exercise asks you to practise calculating minimal polynomials of algebraic elements and looking at the corresponding extensions.

Exercise 2.3

(a) Write down the minimal polynomial for $i = \sqrt{-1}$ over $\mathbb{Q}$.

(b) Let $E = \mathbb{Z}_5(\alpha)$ where $\alpha^2 = 2$ and $F = \mathbb{Z}_5(\beta)$ where $\beta^2 = 3$.

(i) Write down the minimal polynomials for α and β.

(ii) Show that $\alpha \in \mathbb{Z}_5(\beta)$ so that $\mathbb{Z}_5(\alpha) = \mathbb{Z}_5(\beta)$.

Exercise 2.3(b) demonstrates that the same extension can be obtained by adjoining elements with different minimal polynomials.

Exercise 2.4

Let α be a root of the polynomial $x^3 - 3x - 1$. Find the degree of the extension $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$.

Exercise 2.5

Let E be an extension of a field F and let $h(x)$ be an irreducible polynomial over F, and let $\alpha, \beta \in E$ be roots of $h(x)$. Show that $F(\alpha) \cong F(\beta)$.

It is not always the case that algebraic extensions of the same degree are isomorphic. The next worked exercise illustrates this fact.

Worked Exercise 2.24

Show that the fields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{5})$ are not isomorphic.

Solution

Suppose, by way of contradiction, that $\phi\colon\ \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{5})$ is an isomorphism. Since both fields are subfields of $\mathbb{C}$, we know from Theorem 1.14 that $\phi(x) = x$ for all $x \in \mathbb{Q}$. In particular $\phi(2) = 2$.

Now the image of the element $\sqrt{2} \in \mathbb{Q}(\sqrt{2})$ under ϕ must have the form $r + s\sqrt{5}$ for some $r, s \in \mathbb{Q}$. Applying ϕ to both sides of the equation $2 = \sqrt{2}\sqrt{2}$ gives

$$2 = \phi(2) = \phi(\sqrt{2}\sqrt{2}) = \phi(\sqrt{2})\phi(\sqrt{2}) = (r + s\sqrt{5})^2,$$

which implies that $2 = r^2 + 5s^2 + 2rs\sqrt{5}$.

Since the set $\{1, \sqrt{5}\}$ is a basis for $\mathbb{Q}(\sqrt{5})$, we can compare coefficients and deduce that $r^2 + 5s^2 = 2$ and $rs = 0$. If $rs = 0$ then either $s = 0$, so that $2 = r^2$, which contradicts the fact that $r \in \mathbb{Q}$, or $r = 0$, so that $2 = 5s^2$, contrary to the fact that $s \in \mathbb{Q}$.

We conclude that the fields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{5})$ are not isomorphic.

The next exercise elaborates on Worked Exercise 2.24 to illustrate a case where two extensions of a finite field are isomorphic.

Worked Exercise 2.25

ξ is the Greek letter 'xi'.

Consider the algebraic extensions $\mathbb{Z}_3(\xi)$, where $\xi^2 + \xi + 2 = 0$, and $\mathbb{Z}_3(i)$, where $i^2 + 1 = 0$.

(a) Write down the successive powers of ξ in terms of the basis $\{1, \xi\}$.

(b) Find an isomorphism between the two fields.

Solution

(a) The equation $\xi^2 + \xi + 2 = 0$ and arithmetic in $\mathbb{Z}_3(\xi)$ give $\xi^2 = 1 - \xi$, $\xi^3 = -1 - \xi$, $\xi^4 = -1$, $\xi^5 = -\xi$, $\xi^6 = \xi - 1$, $\xi^7 = 1 + \xi$, $\xi^8 = 1$. Since $\{1, \xi\}$ is a basis and $|\mathbb{Z}_3| = 3$, we know that $|\mathbb{Z}_3(\xi)| = 9$, so the multiplicative group of non-zero elements of $\mathbb{Z}_3(\xi)$ has 8 elements. It follows that ξ generates this group.

(b) We give three possible solutions.

(i) Suppose ϕ is a map from $\mathbb{Z}_3(i)$ to $\mathbb{Z}_3(\xi)$ such that $\phi(i) = a + b\xi$. Applying ϕ to $-1 = i^2$ gives

$$\begin{aligned} -1 = \phi(i^2) = (\phi(i))^2 &= (a + b\xi)^2 \\ &= a^2 + 2ab\xi + b^2\xi^2 \\ &= a^2 + 2ab\xi + b^2(1 - \xi) \\ &= a^2 + b^2 + (2a - b)b\xi, \end{aligned}$$

which implies $a^2 + b^2 = -1$ and $(2a - b)b = 0$.

From the second of these equations we deduce that either $b = 0$ or $b = -a$. If $b = 0$ the first equation reduces to $a^2 = -1$, which has no solutions mod 3. If $b = -a$ the first equation reduces to $a^2 = 1$, which has two solutions $a = 1$ and $a = -1$. We take the second of these, so $b = 1 = -a$, and define $\phi(i) = -1 + \xi$.

Thus, we have obtained a map from $\mathbb{Z}_3(i)$ to $\mathbb{Z}_3(\xi)$ defined by $\phi(a + bi) = a + b(-1 + \xi) = a - b + b\xi$. We check that ϕ is a field isomorphism. The map is an additive homomorphism, because

$$\begin{aligned} \phi(a + bi + c + di) &= \phi((a + c) + (b + d)i) \\ &= a + c - (b + d) + (b + d)\xi \\ &= (a - b) + b\xi + (c - d) + d\xi \\ &= \phi(a + bi) + \phi(c + di). \end{aligned}$$

We show that the map ϕ is a multiplicative homomorphism. On the one hand,

$$\begin{aligned} \phi((a + bi)(c + di)) &= \phi(ac - bd + (ad + bc)i) \\ &= (ac - bd - ad - bc) + (ad + bc)\xi. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} \phi(a + bi)\phi(c + di) &= (a - b + b\xi)(c - d + d\xi) \\ &= (a - b)(c - d) + (b(c - d) \\ &\quad + d(a - b))\xi + bd\xi^2 \\ &= (a - b)(c - d) + bd + (b(c - d) \\ &\quad + d(a - b) - bd)\xi \\ &= ac - ad - bc - bd + (bc + ad)\xi. \end{aligned}$$

Therefore

$$\phi((a + bi)(c + di)) = \phi(a + bi)\phi(c + di).$$

Since ϕ is onto and the fields have the same cardinality, ϕ is an isomorphism.

(ii) For our second solution, we observe that any isomorphism between the two fields must take an element of multiplicative order 8 to an element of the same order, so we define $\psi(1+i) = \xi$. This implies that

$$1 + \psi(i) = \xi, \quad \text{or, equivalently,} \quad \psi(i) = \xi - 1, \quad \text{and}$$
$$\psi(a+bi) = a + b(\xi - 1) = a - b + b\xi,$$

obtaining the same candidate for an isomorphism of the fields as in solution (i). The proof is concluded as in the first solution.

(iii) A third solution is obtained by proving that the two extensions are, in fact, equal. We observe that

$$(1-i)^2 + (1-i) + 2 = -2i + (1-i) + 2 \equiv 0 \pmod 3,$$

so that there is a solution of $\xi^2 + \xi + 2 = 0$ in $\mathbb{Z}_3(i)$. This shows that $\mathbb{Z}_3(\xi) \subseteq \mathbb{Z}_3(i)$. Since $[\mathbb{Z}_3(i) : \mathbb{Z}_3] = 2$ and there is no solution to $\xi^2 + \xi + 2 = 0$ in $\mathbb{Z}_3$, this is only possible if $\mathbb{Z}_3(\xi) = \mathbb{Z}_3(i)$. Hence the two extensions are not just isomorphic, but equal.

In Corollary 2.23 you saw that extensions of a field obtained by adjoining a different root of the same irreducible polynomials are isomorphic, no matter which root one chooses. The next example shows that the extensions, although isomorphic, need not be equal to each other.

Example 2.26 *Distinct isomorphic field extensions*

Consider the field $\mathbb{Q}$ and the polynomial $x^3 - 2 \in \mathbb{Q}[x]$. By Theorem 2.19 we know that there exists an extension $\mathbb{Q}(\alpha)$ of $\mathbb{Q}$ with $\alpha^3 - 2 = 0$. Since $x^3 - 2$ is monic and irreducible over $\mathbb{Q}$, we know that $x^3 - 2$ is the minimal polynomial for α. We also know that

$$x^3 - 2 = (x - \sqrt[3]{2})(x - \omega\sqrt[3]{2})(x - \omega^2\sqrt[3]{2}),$$

where $\omega = e^{2\pi i/3}$. Note that $\omega^2 + \omega + 1 = 0$. How is α related to $\sqrt[3]{2}$, $\omega\sqrt[3]{2}$ and $\omega^2\sqrt[3]{2}$?

The answer is that α could, in fact, be any of these three elements, and by Corollary 2.23 we have

$$\mathbb{Q}(\alpha) \cong \mathbb{Q}(\sqrt[3]{2}) \cong \mathbb{Q}(\omega\sqrt[3]{2}) \cong \mathbb{Q}(\omega^2\sqrt[3]{2}) \cong \mathbb{Q}[x]/\langle x^3 - 2\rangle.$$

However, the fact that the three fields are isomorphic does not imply that they are identical. Indeed, we will show this is not the case.

We begin by observing that $\omega \notin \mathbb{R}$ but $\sqrt[3]{2} \in \mathbb{R}$. It follows that, although $\mathbb{Q}(\sqrt[3]{2}) \cong \mathbb{Q}(\omega\sqrt[3]{2})$, the two fields are not identical.

Now suppose that $\mathbb{Q}(\omega\sqrt[3]{2}) = \mathbb{Q}(\omega^2\sqrt[3]{2})$. Multiplying the equation $\omega^2 + \omega + 1 = 0$ by $\sqrt[3]{2}$ and rearranging terms gives

$$\omega^2\sqrt[3]{2} = -\omega\sqrt[3]{2} - \sqrt[3]{2}.$$

This equation shows that if $\omega^2\sqrt[3]{2} \in \mathbb{Q}(\omega\sqrt[3]{2})$, then also $\sqrt[3]{2} \in \mathbb{Q}(\omega\sqrt[3]{2})$. This would mean that $\mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{Q}(\omega\sqrt[3]{2})$ and so, since both extensions have degree 3 over $\mathbb{Q}$ we would have $\mathbb{Q}(\sqrt[3]{2}) = \mathbb{Q}(\omega\sqrt[3]{2})$, which is not the case. Hence the extensions $\mathbb{Q}(\omega\sqrt[3]{2})$ and $\mathbb{Q}(\omega^2\sqrt[3]{2})$ are distinct.

Thus we have shown that $\mathbb{C}$ contains three isomorphic fields that are extensions of degree 3 over $\mathbb{Q}$ and that contain a root of the polynomial $x^3 - 2$. As you will see in Section 4, we can go a step further by noting that $x^2 + x + 1$ is an irreducible polynomial in $\mathbb{Q}(\sqrt[3]{2})[x]$ and by considering a further algebraic extension to obtain $\mathbb{Q}(\sqrt[3]{2}, \omega)$.

The next three exercises offer you a chance to try computations with field extensions.

Exercise 2.6

Prove that the six fields $\mathbb{Z}_{13}(\sqrt{t})$ for $t \in \{\pm 2, \pm 5, \pm 6\}$ are mutually isomorphic.

Exercise 2.7

Show that the fields $\mathbb{Q}(i)$ and $\mathbb{Q}(j)$, where j is a root of $x^2 - 2x + 2$, are isomorphic.

Exercise 2.8

Show that the fields $\mathbb{Q}(\xi)$ with $\xi^2 = -2$ and $\mathbb{Q}(\eta)$, where η satisfies $x^2 - 2x - 2 = 0$, cannot be isomorphic.

η is the Greek letter 'eta'

Exercises 2.7 and 2.8 asked you to work with some algebraic extensions of the field $\mathbb{Q}$ of rational numbers. In this setting, it is quite natural to ask when two extensions of $\mathbb{Q}$, each obtained by adjoining the square root of a rational number, are isomorphic.

Theorem 2.27

Let t_1 and t_2 be positive rational numbers such that both $\sqrt{t_1}$ and $\sqrt{t_2}$ are irrational. Then the fields $\mathbb{Q}(\sqrt{t_1})$ and $\mathbb{Q}(\sqrt{t_2})$ are isomorphic if, and only if, there exists a rational number s such that $t_1 = s^2 t_2$.

Proof We can write the elements of $\mathbb{Q}(\sqrt{t_1})$ and $\mathbb{Q}(\sqrt{t_2})$ as $a + b\sqrt{t_1}$ and $c + d\sqrt{t_2}$ respectively.

Suppose that $\phi : \mathbb{Q}(\sqrt{t_1}) \to \mathbb{Q}(\sqrt{t_2})$ is an isomorphism, and suppose further that $\phi(\sqrt{t_1}) = r + s\sqrt{t_2}$ for some $r, s \in \mathbb{Q}$. Since t_1 is rational, we have

$$t_1 = \phi(t_1) = \phi(\sqrt{t_1}\sqrt{t_1}) = \phi(\sqrt{t_1})\phi(\sqrt{t_1}) = (r + s\sqrt{t_2})^2,$$

which gives

$$t_1 = r^2 + t_2 s^2 + 2rs\sqrt{t_2}.$$

Since t_1 is non-zero, at least one of r, s is non-zero. If $rs \neq 0$, then the last equation gives $\sqrt{t_2} = \frac{t_1 - r^2 - t_2 s^2}{2rs}$, which expresses the irrational number $\sqrt{t_2}$ by a rational number, a contradiction. If $s = 0$, then the equation for t_1 collapses to $t_1 = r^2$, which contradicts the fact that $r \in \mathbb{Q}$. Thus $r = 0$, and the equation for t_1 reduces to $t_1 = s^2 t_2$ as required.

On the other hand, if $t_1 = s^2 t_2$ then $a + b\sqrt{t_1} = a + bs\sqrt{t_2}$ and so $\mathbb{Q}(\sqrt{t_1}) = \mathbb{Q}(\sqrt{t_2})$, that is, the fields $\mathbb{Q}(\sqrt{t_1})$ and $\mathbb{Q}(\sqrt{t_2})$ are not only isomorphic but, in fact, identical. ■

The next exercise illustrates Theorem 2.27.

Exercise 2.9

(a) Show that $\mathbb{Q}(\sqrt{2}) \cong \mathbb{Q}(-\sqrt{2})$.

(b) Show that $\mathbb{Q}(\sqrt{2}) \ncong \mathbb{Q}(\sqrt{-2})$.

In order to achieve a higher level of generality in the theory of field extensions, let us show that the isomorphism of two field extensions obtained by adjoining roots of an irreducible polynomial carries over to the case where the underlying fields are isomorphic (rather than equal). We will need such a general result in Section 4, where we consider fields containing all the roots of a given polynomial.

Theorem 2.28

Let F_1, F_2 be fields. Let $\phi : F_1 \to F_2$ be a field isomorphism and let $\hat{\phi}$ be the homomorphism $\hat{\phi} : F_1[x] \to F_2[x]$ that extends ϕ to $F_1[x]$, so that

$$\hat{\phi}(a_0 + a_1 x + \cdots + a_n x^n) = \phi(a_0) + \phi(a_1)x + \cdots + \phi(a_n)x^n.$$

Let $h_1(x)$ be an irreducible polynomial over F_1 and let $h_2(x) = \hat{\phi}(h_1(x))$ be its (irreducible) image under the mapping $\hat{\phi}$. Let α_1 be a root of $h_1(x)$ and let α_2 be a root of $h_2(x)$. Then the fields $F_1(\alpha_1)$ and $F_2(\alpha_2)$ are isomorphic.

Proof From Corollary 2.23 it follows that $F_1(\alpha_1) \cong F_1[x]/\langle h_1(x)\rangle$ and $F_2(\alpha_2) \cong F_2[x]/\langle h_2(x)\rangle$. Let $\theta : F_2[x] \to F_2[x]/\langle h_2(x)\rangle$ be the natural homomorphism and let σ be the composition $\theta \circ \hat{\phi}$. Then, for every $g(x) \in F_1(x)$ we have $\sigma(g(x)) = \hat{\phi}(g(x)) + \langle h_2(x)\rangle$.

Now,

$$\begin{aligned}\mathrm{Ker}(\sigma) &= \{g(x) : \hat{\phi}(g(x)) \in \langle h_2(x)\rangle\}\\ &= \{g(x) : \hat{\phi}(g(x)) = a(x)h_2(x) \text{ for some } a(x) \in F_2[x]\}\\ &= \{g(x) : g(x) = (\hat{\phi})^{-1}(a(x)h_2(x)) \text{ for some } a(x) \in F_2[x]\}\\ &= \{g(x) : g(x) = b(x)h_1(x) \text{ for some } b(x) \in F_1[x]\}\\ &= \langle h_1(x)\rangle.\end{aligned}$$

Thus, by the First Isomorphism Theorem,

$$F_1[x]/\langle h_1(x)\rangle \cong \mathrm{Im}\,\sigma = F_2[x]/\langle h_2(x)\rangle$$

and so, by Theorem 2.19, $F_1(\alpha_1) \cong F_2(\alpha_2)$ as required. ■

The final theorem in this subsection enables us to find a basis for extensions obtained by adjoining more than one element. Although the proof looks long, the idea is straightforward – namely, forming a basis from the products of elements of the bases of extensions of lower degree.

Theorem 2.29 *The KLM Theorem for field extensions*

Let K, L and M be fields such that $K \subset L \subset M$ and let the degrees $[M : L]$ and $[L : K]$ be both finite. Then

$$[M : L][L : K] = [M : K].$$

Proof We use the fact that extensions can be regarded as vector spaces over the smaller fields.

Since the degree $[M : L]$ is finite, say, equal to m, the vector space M has a basis $B_M = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ over the field L. That is, the set B_M is linearly independent over L and generates M, which means that every element γ of M can be written as a linear combination

$$\gamma = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_m\alpha_m \tag{1}$$

where $c_j \in L$, $1 \leq j \leq m$.

By the same token, since the degree $[L : K]$ is finite, say equal to n, the vector space L has a basis $B_L = \{\beta_1, \beta_2, \ldots, \beta_n\}$ over the field K. As above, this means that the set B_L is linearly independent over the field K and generates L, that is, every element of L can be written as a linear combination of elements in B_L. In particular, this is true for any of the elements $c_1, c_2, \ldots, c_m$. That is, for every $j = 1, 2, \ldots, m$ we have

$$c_j = c_{j1}\beta_1 + c_{j2}\beta_2 + \cdots + c_{jn}\beta_n \tag{2}$$

where $c_{jk} \in K$ for all j, k such that $1 \leq j \leq m$ and $1 \leq k \leq n$.

Substituting the equations (2) in (1) for each $k = 1, 2, \ldots, n$ gives

$$\begin{aligned}\gamma = \quad & c_{11}\alpha_1\beta_1 + c_{12}\alpha_1\beta_2 + \cdots + c_{1n}\alpha_1\beta_n \\ & + c_{21}\alpha_2\beta_1 + c_{22}\alpha_2\beta_2 + \cdots + c_{2n}\alpha_2\beta_n \\ & + \cdots \\ & + c_{m1}\alpha_m\beta_1 + c_{m2}\alpha_m\beta_2 + \cdots + c_{mn}\alpha_m\beta_n\end{aligned}$$

where $c_{jk} \in K$, $1 \leq j \leq m$, $1 \leq k \leq n$.

This shows that the field M as a vector space over K is generated by the mn elements $\alpha_j\beta_k$, where $1 \leq j \leq m$ and $1 \leq k \leq n$.

We conclude the proof by showing that the mn elements $\alpha_j\beta_k$ with $1 \leq j \leq m$ and $1 \leq k \leq n$ are linearly independent over K. Suppose that

$$\begin{aligned}0 = \quad & d_{11}\alpha_1\beta_1 + d_{12}\alpha_1\beta_2 + \cdots + d_{1n}\alpha_1\beta_n \\ & + d_{21}\alpha_2\beta_1 + d_{22}\alpha_2\beta_2 + \cdots + d_{2n}\alpha_2\beta_n \\ & + \cdots \\ & + d_{m1}\alpha_m\beta_1 + d_{m2}\alpha_m\beta_2 + \cdots + d_{mn}\alpha_m\beta_n\end{aligned}$$

for some $d_{jk} \in K$, $1 \leq j \leq m$, $1 \leq k \leq n$. Re-grouping the elements gives

$$\begin{aligned}0 = \quad & (d_{11}\beta_1 + d_{12}\beta_2 + \cdots + d_{1n}\beta_n)\alpha_1 \\ & + (d_{21}\beta_1 + d_{22}\beta_2 + \cdots + d_{2n}\beta_n)\alpha_2 \\ & + \cdots \\ & + (d_{m1}\beta_1 + d_{m2}\beta_2 + \cdots + d_{mn}\beta_n)\alpha_m.\end{aligned}$$

This equality expresses 0 as a linear combination of vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ in M, with coefficients in L. Since the set $\alpha_1, \alpha_2, \ldots, \alpha_m$ is linearly independent over L, we conclude that

$$\begin{aligned}0 &= d_{11}\beta_1 + d_{12}\beta_2 + \cdots + d_{1n}\beta_n \\ 0 &= d_{21}\beta_1 + d_{22}\beta_2 + \cdots + d_{2n}\beta_n \\ &\cdots \\ 0 &= d_{m1}\beta_1 + d_{m2}\beta_2 + \cdots + d_{mn}\beta_n.\end{aligned}$$

For each j such that $1 \leq j \leq m$, the jth equality above expresses 0 as a linear combination of vectors $\beta_1, \beta_2, \ldots, \beta_n$ in L with coefficients d_{jk} in K. Since the vectors $\beta_1, \beta_2, \ldots, \beta_n$ are linearly independent over K, we conclude that $d_{jk} = 0$ for all j, k such that $1 \leq j \leq m$ and $1 \leq k \leq n$.

Thus, the set $\{\alpha_j\beta_k : \ 1 \leq j \leq m, \ 1 \leq k \leq n\}$ is indeed a basis for M. It follows that the degree of M over K is mn, and hence $[M : L][L : K] = [M : K]$, as claimed. ■

Because of the symbols appearing in its formulation, this result is sometimes nicknamed the 'Royal Dutch Airlines' or 'KLM' Theorem.

Exercise 2.10

Let $L = \mathbb{Q}(\sqrt{2})$. Show that $x^2 - 3$ is irreducible over L, that is, show that there are no $a, b \in \mathbb{Q}$ such that $(a + b\sqrt{2})^2 = 3$. What is $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}]$?

2.4 Transcendental extensions

Subsection 2.3 presented a selection of basic results concerning the structure of algebraic extensions. This subsection focuses on the structure of transcendental extensions. Recall that if E is an extension of a field F, an element $\alpha \in E \backslash F$ is said to be transcendental over F if α is not a root of any polynomial over F.

Theorem 2.30

Let E be an extension of a field F and let $\alpha \in E$ be a transcendental element over F. Then the smallest subfield $F(\alpha)$ of E containing F and α has the form

$$F(\alpha) = \{f(\alpha)(g(\alpha))^{-1} : \ f(x), g(x) \in F[x], \ g(x) \neq 0\}.$$

That is, $F(\alpha)$ is isomorphic to the field of fractions of the integral domain $F[x]$.

Proof The evaluation at α of a polynomial $g(x) \in F[x]$, that is, the value $g(\alpha)$, belongs to $F(\alpha)$ since it consists of sums and products of elements of F and powers of α. Note that, since α is transcendental, $g(\alpha)$ is non-zero if $g(x)$ is non-zero, and any non-zero element $g(\alpha) \in F(\alpha)$ has a multiplicative inverse $(g(\alpha))^{-1}$.

Also, for any $f(x) \in F[x]$ one may form the product $f(\alpha)(g(\alpha))^{-1}$.

Thus, if α is transcendental, the smallest subfield $F(\alpha)$ of F that contains both F and α must, at the very least, contain all the elements of the form $f(\alpha)(g(\alpha))^{-1}$, where $f(x) \in F[x]$ is any polynomial and $g(x) \in F[x]$ is a non-zero polynomial.

We now show that $F(\alpha)$ contains no other elements. In order to do so, it is sufficient to show that the set

$$F_\alpha = \{f(\alpha)(g(\alpha))^{-1} : \ f(x), g(x) \in F[x], \ g(x) \neq 0\}$$

is a subfield of E. We use the subfield criterion given in Book C, Chapter 11, Lemma 1.6.

For closure of addition we have:

$$\begin{aligned}
&f(\alpha)(g(\alpha))^{-1} + f_1(\alpha)(g_1(\alpha))^{-1} \\
&\quad = f(\alpha)(g(\alpha))^{-1} g_1(\alpha)(g_1(\alpha))^{-1} + f_1(\alpha)(g_1(\alpha))^{-1} g(\alpha)(g(\alpha))^{-1} \\
&\quad = f(\alpha) g_1(\alpha)(g(\alpha))^{-1}(g_1(\alpha))^{-1} + f_1(\alpha) g(\alpha)(g_1(\alpha))^{-1}(g(\alpha))^{-1} \\
&\quad = [f(\alpha) g_1(\alpha) + f_1(\alpha) g(\alpha)][g(\alpha)^{-1} g_1(\alpha)^{-1}] \in F_\alpha \\
&\quad = [f(\alpha) g_1(\alpha) + f_1(\alpha) g(\alpha)](g(\alpha) g_1(\alpha))^{-1} \in F_\alpha,
\end{aligned}$$

where this calculation comes from the familiar method for adding fractions. (Note the ample use of commutativity in these calculations.)

For closure of multiplication we have:

$$f(\alpha)(g(\alpha))^{-1} \times f_1(\alpha)(g_1(\alpha))^{-1} = f(\alpha) f_1(\alpha)[g(\alpha) g_1(\alpha)]^{-1} \in F_\alpha.$$

It is clear that $1 \in F_\alpha$. Moreover, if $f(\alpha)(g(\alpha))^{-1} \in F_\alpha$ and $f(x)$ is non-zero, then the multiplicative inverse $g(\alpha)(f(\alpha))^{-1}$ is also in F_α. Hence F_α is a subfield of E.

Since F_α is a field containing both F and α, it follows that $F(\alpha) \subseteq F_\alpha$. On the other hand, all the elements of F_α must be contained in $F(\alpha)$, and so $F_\alpha = F(\alpha)$.

The isomorphism with the field of fractions of polynomials over F follows from Definition 1.6 of Chapter 17. ■

We have just shown that a given transcendental extension takes a particular form. We now show, perhaps in a surprisingly brief way, that every field has a transcendental extension.

Theorem 2.31

For every field L, the field of fractions of the integral domain $L[x]$ is an extension of L that contains a transcendental element over L.

Proof Let L be an arbitrary field. Consider the ring $L[x]$ of all polynomials in the indeterminate x with coefficients in L. We know from Book C, Chapter 11, Lemma 2.8 that $L[x]$ is an integral domain. We may therefore construct the corresponding field of fractions and denote it by F.

By this construction, elements of F can be identified with fractions of the form $\frac{f(x)}{g(x)}$, where $f(x), g(x) \in L[x]$ and $g(x)$ is a non-zero polynomial.

We now claim that the polynomial $t(x) = x$ is a transcendental element of the field F over L. Indeed, suppose that $t(x)$ is a root of some non-zero polynomial $f(x) \in L[x]$ in some extension E of L. Then substituting the element $t(x) \in E$ for x into $f(x)$ would give zero, that is, $f(t(x)) = 0$ in E.

Since $t(x)$ is an element of both the fields E and F, it is also an element of the field $E \cap F$. Therefore in $E \cap F$ we have $f(t(x)) = 0$. But since $t(x) = x$, this would mean that $f(x) = 0$ in $E \cap F$; in particular, $f(x) = 0$ in F, that is, f would be a zero polynomial in $L[x]$, contrary to our assumption about $f(x)$. Thus, the element $x \in F$ is transcendental over L, as claimed. ■

Theorem 2.30 shows in particular that the extension of a field by a transcendental element is isomorphic to the field of fractions of the polynomial ring over the field. This result gives the following corollary.

Corollary 2.32

Let E be a field extension of a field F. Suppose that $\alpha, \beta \in E$ and both α and β are transcendental elements over F. Then $F(\alpha) \cong F(\beta)$.

Proof This follows directly from Theorem 2.30 since both $F(\alpha)$ and $F(\beta)$ are isomorphic to the field of fractions of $F(x)$. ■

3 Finite fields

Finite fields are an important area of study and have significant applications in digital information exchange through coding theory and cryptography.

This section begins with an investigation of the structure of the multiplicative group whose elements are the non-zero elements of a finite field. For a field F, this group is usually denoted by F^*.

3.1 The multiplicative group of a finite field

We prove that the multiplicative group of a finite field is cyclic by finding an element that has the same order as that of the group. This involves a counting argument about elements of particular orders. Recall that, by Lagrange's Theorem (Book B, Chapter 5, Theorem 2.16), the order of an element must divide the order of the group. The next exercise identifies elements of a given order within the multiplicative group.

Exercise 3.1

Let F be a finite field of order p^n for some prime p. Further, let d be a positive divisor of $p^n - 1$ and let u be an element of order d in $F^* = F\backslash\{0\}$.

The fact that every finite field has order of the form p^n follows from Theorem 2.9.

(a) Show that any element of the set $U = \{u, u^2, \ldots, u^{d-1}, 1\}$ is a root of the polynomial $x^d - 1$.

(b) Show that for $u \in U$ we have $u^i \neq u^j$ if $0 \leq i < j < d$.

In the counting argument that we are about to develop, we use the Euler ϕ-function that gives, for a positive integer $m \geq 2$, the number of positive integers smaller than m and relatively prime to m (with $\phi(1) = 1$). You encountered this function in Book C, Chapter 9, Section 3.

Lemma 3.1

Let F be a finite field of order p^n for some prime p and let d be a positive divisor of $p^n - 1$. Let a_d be the number of elements of $F^* = F\backslash\{0\}$ that have multiplicative order equal to d. If $a_d > 0$ then $a_d = \phi(d)$ where $\phi(d)$ is the value of the Euler ϕ-function at d.

There are elements of U of order dividing d but less than d. For example, 1 is such an element.

Proof If $a_d > 0$, there exists an element $u \in F^*$ of order d. By Exercise 3.1, $U = \{u, u^2, \ldots, u^{d-1}, 1\}$ is a set of d distinct roots of the polynomial $x^d - 1$ over F. By Proposition 4.3 of Book C, Chapter 11, the polynomial $x^d - 1$ has at most d roots in F. Thus, the set U contains all the roots of $x^d - 1$ in F, and hence also in F^*. It follows that all the elements of F^* of order d must lie in the set U.

The elements of U that have order *equal* to d (and not just a proper divisor of d) are the powers u^j of u such that $1 \leq j \leq d$ and j is relatively prime to d. The number of such exponents j is the value $\phi(d)$ of the Euler ϕ-function. The result follows. ■

Exercise 3.2

Let n be a positive integer and let d be a positive divisor of n. Show that all the elements of order d in the additive group $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ have the form sr for $r = \frac{n}{d}$, where $1 \leq s \leq d$ and s and d are relatively prime. Deduce that the number of elements of order d in the cyclic group $(\mathbb{Z}_n, +)$ is equal to the value $\phi(d)$ of the Euler ϕ-function.

We need one more auxiliary result.

Lemma 3.2

Let $q = p^n$ for some prime p and let

$$A = \{d : 0 < d < q \text{ and } d \text{ divides } q - 1\}.$$

Then $q - 1 = \sum_{d \in A} \phi(d)$, where $\phi(d)$ is the Euler ϕ-function.

Proof By Exercise 3.2, for each positive divisor d of $q - 1$ the group $\mathbb{Z}_{q-1}$ contains exactly $\phi(d)$ elements of order equal to d. Since every element of $\mathbb{Z}_{q-1}$ has an order that divides $q - 1$, we have

$$q - 1 = \sum_{d|(q-1)} \phi(d).$$

■

We are now ready to prove our main result about multiplicative groups of finite fields.

Theorem 3.3 *Multiplicative group of a finite field*

The multiplicative group of every finite field is cyclic.

Proof Let F be a finite field of order $q = p^n$ for some prime p. We know that its multiplicative group $F^* = F\backslash\{0\}$ is abelian and has order $q-1$. We need to show that this group is, in fact, cyclic.

Firstly, for any positive divisor d of $q-1$ let a_d be the number of elements of F^* of multiplicative order equal to d. Since every element of F^* has some order and this order is a positive divisor of $q-1$, we have

$$\sum_{d|(q-1)} a_d = q-1.$$

Secondly, by Lemma 3.1 if $a_d > 0$ then $a_d = \phi(d)$.

Thirdly, by Lemma 3.2 we have $q-1 = \sum_{d|(q-1)} \phi(d)$.

These three conditions hold simultaneously only when $a_d = \phi(d)$ for every positive divisor d of $q-1$, that is, it is impossible to have $a_d = 0$ if d divides $q-1$. In particular, for $d = q-1$ we have $a_{q-1} = \phi(q-1)$. This means that the group F^*, of order $q-1$, contains an element of order equal to $q-1$ (in fact, our argument shows that F^* contains $\phi(q-1)$ such elements). But the existence of an element of order $q-1 = |F^*|$ means that the group F^* is cyclic, since such an element generates F^*. ■

In the theory of finite fields, generators of the multiplicative groups play such an important role that they have a special name.

Definition 3.4 *Primitive element*

Any generator of the multiplicative group of a finite field F is called a **primitive element** of F.

With the help of Euler's ϕ-function we are able to determine the number of primitive elements in any finite field.

Corollary 3.5

For any prime p, let F be a field of order $q = p^n$.

(a) Let ξ be a primitive element of F. Then

$$\{\xi^j : 1 \le j < q-1,\ j \text{ and } q-1 \text{ are relatively prime}\}$$

is the set of all the primitive elements of F.

(b) The number of primitive elements of a finite field of order p^n is equal to $\phi(p^n - 1)$.

Proof

(a) Since, by definition, ξ is a generator of the multiplicative group F^* of F, we have $F^* = \{1, \xi, \xi^2, \ldots, \xi^{q-2}\}$, and hence the complete list of the elements of F is $\{0, 1, \xi, \ldots, \xi^j, \ldots, \xi^{q-2}\}$. Then the map $\xi^j \mapsto j$ from the multiplicative group F^* onto the additive group $\mathbb{Z}_{q-1} = \{0, 1, \ldots, j, \ldots, q-2\}$ is an isomorphism. We know that the generators of $\mathbb{Z}_{q-1}$ are the elements j such that j and $q-1$ are relatively prime. The above isomorphism gives the required result.

(b) This follows from (a) and from the definition of $\phi(p^n - 1)$. ■

The next five exercises should help you become familiar with various aspects of calculations in finite fields.

Exercise 3.3

Let $\mathbb{Z}_3(i)$ be the extension field of $\mathbb{Z}_3$ by i where $i^2 + 1 = 0$.

(a) Write down the successive powers of $1 + i$.

(b) Find the primitive elements of the field $\mathbb{Z}_3(i)$.

Exercise 3.4

Show that 2 is a primitive element in $\mathbb{Z}_5$, $\mathbb{Z}_{11}$, $\mathbb{Z}_{13}$ and $\mathbb{Z}_{19}$.

Exercise 3.5

Show that 3 is a primitive element in $\mathbb{Z}_7$, $\mathbb{Z}_{17}$ and $\mathbb{Z}_{19}$.

Exercise 3.6

Express every non-zero element of the field $\mathbb{Z}_{11}$ in the form 2^n, $1 \leq n \leq 10$.

(a) Write down the non-zero elements of $\mathbb{Z}_{11}$ that are squares. For which values of $a \in \mathbb{Z}_{11}^*$ is the polynomial $t^2 - a$ irreducible?

(b) Write down the non-zero elements of $\mathbb{Z}_{11}$ that are cubes. For which values of $a \in \mathbb{Z}_{11}^*$ is the polynomial $t^3 - a$ irreducible?

(c) Find all primitive elements of $\mathbb{Z}_{11}$ and express them as powers of 2.

Exercise 3.7

Express every non-zero element in the field $\mathbb{Z}_{17}$ in the form 3^n, $1 \leq n \leq 16$.

(a) Write down the non-zero elements of $\mathbb{Z}_{17}$ that are squares. For which values of $a \in \mathbb{Z}_{17}^*$ is the polynomial $t^2 - a$ irreducible?

(b) Write down the non-zero elements of $\mathbb{Z}_{17}$ that are cubes. For which values of $a \in \mathbb{Z}_{17}^*$ is the polynomial $t^3 - a$ irreducible?

(c) Find all the primitive elements of $\mathbb{Z}_{17}$ and express them as powers of 3.

3.2 Roots and inverses in the multiplicative group

Primitive elements in finite fields are those of largest order in the (cyclic) multiplicative group of the field. The next definition offers a different viewpoint on the orders of the elements of the multiplicative group of a general (not necessarily finite) field.

Definition 3.6 *Roots of unity*

Let F be a field. Then any root of the polynomial $x^k - 1$ in F is known as a kth **root of unity**. Those elements $u \in F$ that have multiplicative order k in the multiplicative group F^* are called **primitive** kth roots of unity.

Since Definition 3.6 applies to arbitrary fields, you may want to know what it says about the field $\mathbb{C}$ of complex numbers that you are familiar with.

Example 3.7 *Roots of unity in $\mathbb{C}$*

If F is the field $\mathbb{C}$ of complex numbers, then the kth roots of unity are precisely the k complex numbers of the form $\cos(\frac{2\pi m}{k}) + i\sin(\frac{2\pi m}{k})$, where $0 \leq m \leq k-1$. Observe that if $\eta_k = \cos(\frac{2\pi}{k}) + i\sin(\frac{2\pi}{k})$, then, by the well-known de Moivre formula, $\eta_k^m = \cos(\frac{2\pi m}{k}) + i\sin(\frac{2\pi m}{k})$ for all m. Hence the set of all complex kth roots of unity is

$$\{\eta_k^m \ : \ 0 \leq m \leq k-1\}.$$

The set of primitive complex kth roots of unity contains all the elements η_k^m for which k and m are relatively prime. That is, there are precisely $\phi(k)$ primitive complex kth roots of unity, where $\phi(k)$ is the value of Euler's ϕ-function at k.

We now return to finite fields. The next theorem gives a simple criterion for a finite field to contain a primitive kth root of unity.

Theorem 3.8

A finite field F of order a prime power $q = p^n$ contains a primitive kth root of unity if, and only if, k is a divisor of $p^n - 1$. In this case, F contains precisely $\phi(k)$ primitive kth roots of unity.

Proof It is sufficient to note that the existence of a primitive kth root of unity in F is equivalent to the existence of an element of order equal to k in the multiplicative group, F^*, of the field. By Theorem 3.3, the multiplicative group of F is cyclic, and hence isomorphic to the additive group $(\mathbb{Z}_{q-1}, +)$. By Lagrange's Theorem, if this group contains an element of order k, then k is a divisor of $q-1$. Conversely, if k divides $q-1$, then by the result of Exercise 3.2, all elements of order exactly k in the additive group $\mathbb{Z}_{q-1} = \{0, 1, 2, \ldots, q-1\}$ have the form $j(q-1)/k$, where j and k are relatively prime. The result follows. ■

As an application of Theorem 3.8 we prove a useful observation about the existence of a square root (that is, the 2nd root) of -1 in a finite field.

Proposition 3.9

In a finite field F of order p^n, the element -1 has no square root in F if, and only if, $p \equiv 3 \pmod 4$ and n is odd.

Proof The square roots of 1 in F are ± 1 and so any primitive 4th root of unity must be a square root of -1. Hence $-1 \in F$ has a square root in F if, and only if, F has a primitive 4th root of unity.

If $p \equiv 3 \pmod 4$ and $n = 2k+1$, then $p^n = p^{2k+1} = (p^2)^k p \equiv 3 \pmod 4$, since $p^2 \equiv 1 \pmod 4$, and so $p^n - 1 \equiv 2 \pmod 4$. Thus 4 does not divide $p^n - 1$ and so, by Theorem 3.8, F does not have a primitive 4th root of unity. Thus $-1 \in F$ has no square root in F if $p \equiv 3 \pmod 4$.

Now suppose conversely that either $p \not\equiv 3 \pmod 4$ or that n is even.

Firstly, if $p = 2$, then $-1 = 1$ and so -1 has a square root in F, namely, -1 itself.

Secondly, let p be odd and let $p \equiv 1 \pmod 4$. Then $p^n \equiv 1 \pmod 4$ and so $p^n - 1$ is divisible by 4. Then by Theorem 3.8, F has a primitive 4th root of unity, which will be a square root of -1.

Finally, suppose that $p \equiv 3 \pmod 4$ and n is even. Then $p^n \equiv 1 \pmod 4$ and so $p^n - 1$ is divisible by 4. Then by Theorem 3.8, F has a primitive 4th root of unity, which will be a square root of -1. We have shown that in each case F has a square root of -1. The result follows. ■

Until this point you have only encountered finite fields of prime order p, namely, $\mathbb{Z}_p$, or GF_p in an alternative notation. The next exercise asks you to describe an infinite family of fields of order p^2.

Exercise 3.8

Let p be a prime such that $p \equiv 3 \pmod 4$. With the help of Theorem 2.19 and Proposition 3.9, construct a field of order p^2.

You have seen that for a prime p and irreducible polynomial $h(x)$ in $\mathbb{Z}_p[x]$ of degree n, the field $\mathbb{Z}_p[x]/\langle h(x)\rangle$ has order p^n. You also know that the field $\mathbb{Z}_p[x]/\langle h(x)\rangle$ can be represented in the form $\mathbb{Z}_p(\alpha)$ by adjoining a root of the polynomial $h(x)$, that is, an element α such that $h(\alpha) = 0$. This subsection concludes with two worked exercises that illustrate two techniques for finding the multiplicative inverse of a particular element expressed as a polynomial in α.

Worked Exercise 3.10

Let $K = \mathbb{Z}_3(\xi)$ where $\xi^3 = \xi + 1$ (note that the polynomial $x^3 - x - 1$ is irreducible over $\mathbb{Z}_3$, so K is a field).

(a) Find the multiplicative inverse of $2\xi + 1$ in K.

(b) Find the multiplicative inverse of $\xi^2 + 1$ in K.

Solution

(a) We will try to divide $2\xi + 1$ into $1 = \xi^3 - \xi = \xi^3 + 2\xi$ in $\mathbb{Z}_3(\xi)$ (since we are working modulo 3, we have $-2 = 1$ and $-1 = 2$).

$$
\begin{array}{rl}
 & 2\xi^2 + 2\xi \\
\hline
2\xi + 1 \;) & \xi^3 + 2\xi \\
 & \xi^3 + 2\xi^2 \\
\hline
 & \xi^2 + 2\xi \\
 & \xi^2 + 2\xi \\
\hline
\end{array}
$$

We find that

$$(2\xi + 1)(2\xi^2 + 2\xi) = \xi^3 - \xi = 1,$$

so that

$$(2\xi + 1)^{-1} = 2\xi^2 + 2\xi\,.$$

(b) In this instance a single long division will be insufficient. We illustrate a general way of finding multiplicative inverses, which is to multiply the given element, in our case $\xi^2 + 1$, by an arbitrary element of K. The latter can be written in the form $a\xi^2 + b\xi + c$ for $a, b, c \in \mathbb{Z}_3$. To simplify the result, we use $\xi^3 = \xi + 1$ and $\xi^4 = \xi^2 + \xi$:

$$
\begin{aligned}
(\xi^2 + 1)(a\xi^2 + b\xi + c) &= a\xi^4 + b\xi^3 + (a + c)\xi^2 + b\xi + c \\
&= a(\xi^2 + \xi) + b(\xi + 1) + (a + c)\xi^2 + b\xi + c \\
&= (2a + c)\xi^2 + (a + 2b)\xi + b + c.
\end{aligned}
$$

We can then determine the elements a, b, c for which

$$(2a + c)\xi^2 + (a + 2b)\xi + b + c = 1$$

by comparing coefficients. Again, we are working over $\mathbb{Z}_3$, and so to simplify things we replace -1 by 2 (and -2 by 1). This gives the system of equations:

$$
\begin{aligned}
2a + c &= 0 \\
a + 2b &= 0 \\
b + c &= 1.
\end{aligned}
$$

Subtracting the third equation from the first gives $b + a = 1$, and combining this with the second equation gives $b = 2$.

Hence $a = b = c = 2$ and the multiplicative inverse of $\xi^2 + 1$ is $2\xi^2 + 2\xi + 2$ (you can check this by multiplying out $(\xi^2 + 1)(2\xi^2 + 2\xi + 2)$).

Worked Exercise 3.11

Find the multiplicative inverse of the following elements in $K = \mathbb{Z}_3(\xi)$ where $\xi^3 = \xi + 1$.

(a) ξ

(b) $\xi + 1$

(c) $\xi^2 + 2\xi + 1$

Solution

(a) We have $1 = \xi^3 - \xi = \xi(\xi^2 - 1)$ in $\mathbb{Z}_3(\xi)$, which implies that $\xi^{-1} = \xi^2 + 2$ as $-1 \equiv 2 \pmod 3$. (If you used long division or compared coefficients you will have got the same answer but with rather more work.)

(b) We divide $\xi + 1$ into $1 = \xi^3 - \xi = \xi^3 + 2\xi$ in $\mathbb{Z}_3(\xi)$.

$$\begin{array}{rl} & \xi^2 + 2\xi \\ \xi + 1 \;) & \xi^3 \quad + \quad + 2\xi \\ & \xi^3 + \xi^2 \\ \hline & \quad 2\xi^2 + 2\xi \\ & \quad 2\xi^2 + 2\xi \\ \hline \end{array}$$

and find that $(\xi + 1)^{-1} = \xi^2 + 2\xi$.

(c) In this instance one long division is again insufficient and we use the more general method. Since every element of K can be written in the form $a\xi^2 + b\xi + c$ with $a, b, c \in K$, we write

$$\begin{aligned} 1 &= (\xi^2 + 2\xi + 1)(a\xi^2 + b\xi + c) \\ &= a\xi^4 + (2a + b)\xi^3 + (a + 2b + c)\xi^2 + (b + 2c)\xi + c \\ &= a(\xi^2 + \xi) + (2a + b)(\xi + 1) + (a + 2b + c)\xi^2 + (b + 2c)\xi + c \\ &= (2a + 2b + c)\xi^2 + (2b + 2c)\xi + 2a + b + c, \end{aligned}$$

where we used $\xi^3 = \xi + 1$ and $\xi^4 = \xi^2 + \xi$ to simplify the result.

Comparing coefficients in the resulting equation

$$(2a + 2b + c)\xi^2 + (2b + 2c)\xi + 2a + b + c = 1$$

over $\mathbb{Z}_3$ leaves us with the linear system

$$\begin{aligned} 2a + 2b + c &= 0 \\ 2b + 2c &= 0 \\ 2a + b + c &= 1. \end{aligned}$$

Subtracting the third equation from the first gives $b = 2$, and substituting this in the second equation gives $4 + 2c = 0$ and hence $c = 1$. A further substitution in the first equation then gives $2a + 4 + 1 = 0$, so that $a = 2$.

Hence $a = 2$, $b = 2$, $c = 1$, and so the multiplicative inverse of $\xi^2 + 2\xi + 1$ is $2\xi^2 + 2\xi + 1$.

4 Splitting fields

The theory of field extensions originated in the hunt for solutions to polynomial equations. It is therefore reasonable to consider extensions which contain *all* the roots of a given polynomial. This section begins with a formal definition of extensions of these kind, and then shows that for a given polynomial over a field such an extension always exists and is unique up to isomorphism.

4.1 Existence and uniqueness of splitting fields

Definition 4.1 *Splitting field*

Let $f(x)$ be a polynomial of positive degree over a field F. A **splitting field** of $f(x)$ over F is a field E that has the following properties:

(a) E is an extension of F that contains all the roots of $f(x)$

(b) if L is an extension of F that contains all the roots of $f(x)$, then E is isomorphic to a subfield of L.

Definition 4.1(b) is essentially saying that E is the smallest field that contains all the roots of the polynomial.

We first prove that for a given polynomial of positive degree over a field F, there is indeed a field extension of E that contains all the roots of F.

Theorem 4.2

Let $f(x)$ be a polynomial of positive degree over a field F. Then there exists an extension of F that contains all the roots of $f(x)$.

Proof We prove this result by induction on the degree of $f(x)$.

If $f(x)$ has degree one, then $f(x)$ has its (unique) root in F.

Now suppose that the result holds for all polynomials of degree less than k for $k \geq 1$, and suppose that $f(x)$ is a polynomial of degree k.

Let $f(x) = f_1(x)f_2(x)\cdots f_m(x)$ be a factorisation into irreducible factors over F (the value $m = 1$ corresponds to the case when $f(x)$ itself is irreducible). Set $h(x) = f_1(x)$. By Corollary 2.20, $h(x)$, and hence $f(x)$, has a root in the extension $F_1 = F[x]/\langle h(x)\rangle$ of F.

Let $x_1, x_2, \ldots, x_\ell$ be the collection of all the roots of $f(x)$ that are contained in F_1; we know that this collection has at least one member. Over F_1 we then have

$$f(x) = (x - x_1)(x - x_2)\cdots(x - x_\ell)g(x),$$

where the degree of $g(x)$ is smaller than the degree of $f(x)$. By the induction hypothesis, the polynomial $g(x)$ has all its roots in some extension F_2 of F_1. But then, all the roots of $f(x)$ lie in F_2 and this is also an extension of F. ■

Definition 4.3 *Splitting polynomial*

Let F be a field. We say that a polynomial $f(x) \in F[x]$ **splits** over F if all roots of $f(x)$ lie in F.

In your previous studies, you might have met the Fundamental Theorem of Algebra, which states that any non-constant polynomial in $\mathbb{C}[x]$ has at least one root in $\mathbb{C}$. One way of phrasing this pivotal result is to say that every polynomial in $\mathbb{C}[x]$ splits over $\mathbb{C}$.

For the formulation of our next result, let us recall Definition 2.12: if F is an extension of a field K and $\alpha_1, \alpha_2, \ldots, \alpha_d \in F$, the symbol $K(\alpha_1, \ldots, \alpha_d)$ denotes the smallest subfield of F that contains K and all the elements $\alpha_1, \alpha_2, \ldots, \alpha_d$.

Proposition 4.4

Suppose that $f(x)$ is a polynomial in $K[x]$ for some field K and

1. $f(x)$ splits over an extension F of K
2. $\{\alpha_1, \alpha_2, \ldots, \alpha_d\}$ is the set of all roots of $f(x)$.

Then $K(\alpha_1, \ldots, \alpha_d)$ is a splitting field for $f(x)$ over K and $K(\alpha_1, \ldots, \alpha_d) \subseteq F$.

It may or may not be true that $K(\alpha_1, \ldots, \alpha_d) = F$.

Proof Since $f(x)$ splits over the field F, we have $\{\alpha_1, \alpha_2, \ldots, \alpha_d\} \subseteq F$, and so $K(\alpha_1, \ldots, \alpha_d)$ is a subfield of F. Since $K(\alpha_1, \ldots, \alpha_d)$ is the smallest subfield of F containing K and all the roots of $f(x)$, it is a splitting field for $f(x)$. ■

We may now state and prove the following consequence of our considerations.

Corollary 4.5 *Existence of splitting fields*

Let $f(x)$ be a polynomial of positive degree over a field F. Then there exists a splitting field of $f(x)$ over F.

Proof By Theorem 4.2, there exists an extension of F over which $f(x)$ splits. The result then follows by Proposition 4.4. ■

The next theorem shows that a splitting field for a given polynomial is unique up to isomorphism.

Theorem 4.6 *Uniqueness of splitting fields*

Let $\theta : K \to L$ be an isomorphism of fields and let $\hat{\theta} : K[x] \to L[x]$ be the extension of θ to the polynomial ring $K[x]$ defined by

$$\hat{\theta}(a_0 + a_1x + \cdots + a_nx^n) = \theta(a_0) + \theta(a_1)x + \cdots + \theta(a_n)x^n.$$

Let $f(x) \in K[x]$ have degree $d \geq 1$ and let $g(x) = \hat{\theta}(f(x))$ be its image over L. Let K_S, L_S be fields over which $f(x)$ and $g(x)$, respectively, split. Further, let $A = \{\alpha_1, \alpha_2, \ldots, \alpha_d\}$ be the set of all the roots of $f(x)$ in K_S and let $B = \{\beta_1, \beta_2, \ldots, \beta_d\}$ be the set of all the roots of $g(x)$ in L_S.

Then there is an isomorphism $\phi : K(\alpha_1, \ldots, \alpha_d) \to L(\beta_1, \ldots, \beta_d)$ such that

(a) the restriction $\phi_A : A \to B$ of ϕ to A is a bijection

(b) $\phi(k) = \theta(k)$ for every $k \in K$.

Proof We will prove the result by induction on d, the degree of the polynomial $f(x)$.

If $\deg(f) = 1$ then the polynomials $f(x)$ and $g(x)$ split over K and L, respectively, and the result is immediate from the isomorphism of K and L.

Let the degree of $f(x)$ be d for $d > 1$, and suppose that the result holds for all instances where the degree of the polynomials involved is smaller than d.

Let $f_1(x)$ be an irreducible factor of $f(x)$ over K and let $g_1(x) = \hat{\theta}(f_1(x))$ be the corresponding irreducible factor of $g(x)$ over L. We may assume without loss of generality that α_1 is a root of $f_1(x)$ and β_1 is a root of $g_1(x)$.

By Theorem 2.28, there is an isomorphism θ_1 from the field $K_1 = K(\alpha_1)$ onto the field $L_1 = L(\beta_1)$ such that $\theta_1(\alpha_1) = \beta_1$ and $\theta_1(k) = \theta(k)$ for all $k \in K$.

Since α_1 and β_1 are roots of $f(x)$ and $g(x)$ in the fields K_1 and L_1, the two polynomials factor as $f(x) = (x - \alpha_1)f_2(x)$ and $g(x) = (x - \beta_1)g_2(x)$, where $f_2(x)$ and $g_2(x)$ are polynomials over K_1 and L_1, respectively.

Let $\hat{\theta}_1$ be the extension of θ to $K_1[x]$. Then

$$\hat{\theta}_1(f(x)) = (x - \theta_1(\alpha_1))\hat{\theta}_1(f_2(x)), \text{ that is, } \hat{\theta}_1(f(x)) = (x - \beta_1)\hat{\theta}_1(f_2(x))$$

over the field $L_1 = \theta_1(K_1)$. But the coefficients in $f(x)$ on the left-hand side of this equation all lie in K. Hence $\hat{\theta}_1(f(x)) = \hat{\theta}(f(x)) = g(x)$, and therefore $g(x) = (x - \beta_1)\hat{\theta}_1(f_2(x))$.

Recall also that $g(x) = (x - \beta_1)g_2(x)$ for some polynomial $g_2(x)$ over L_1. Comparing the last two equations gives $g_2(x) = \hat{\theta}_1(f_2(x))$ since $L_1[x]$ is a unique factorisation domain (see Book C, Chapter 11, Theorem 3.1).

Hence $K_1(\alpha_2, \ldots, \alpha_d)$ is the splitting field of $f_2(x)$ over K_1, while $L_1(\beta_2, \ldots, \beta_d)$ is the splitting field of $g_2(x)$ over L_1. In addition, $\theta_1 : K_1 \to L_1$ is a field isomorphism such that $\hat{\theta}_1(f_2(x)) = g_2(x)$.

Since the degree of $f_1(x)$ is smaller than d, we may apply the induction hypothesis. That is, we may assume the existence of a field isomorphism $\phi : K_1(\alpha_2, \ldots, \alpha_d) \to L_1(\beta_2, \ldots, \beta_d)$ such that ϕ maps the set $\{\alpha_2, \ldots, \alpha_d\}$ onto the set $\{\beta_2, \ldots, \beta_d\}$, and $\phi(z) = \theta_1(z)$ for all $z \in K_1$.

However, we saw above that $\theta_1(\alpha_1) = \beta_1$ and $\theta_1(k) = \theta(k)$ for each $k \in K$. This implies that $\phi(\alpha_1) = \beta_1$ and $\phi(k) = \theta(k)$ for all $k \in K$. In other words, ϕ is also an isomorphism from the field $K(\alpha_1, \alpha_2, \ldots, \alpha_d)$ onto the field $L(\beta_1, \beta_2, \ldots, \beta_d)$ with the required properties. The result now follows by induction. ■

The next corollary gives a more descriptive statement of Theorem 4.6.

Corollary 4.7

Let $f(x)$ be a polynomial over a field F. Then all splitting fields for $f(x)$ are mutually isomorphic.

The next two worked exercises invite you to practise with splitting fields for quadratic polynomials, that is, polynomials of degree 2.

Worked Exercise 4.8

Show that if $h(x)$ is an irreducible quadratic polynomial in $F[x]$, then $E = F[x]/\langle h(x)\rangle$ is a splitting field for $h(x)$ over F.

Hint: show that $h(x)$ factorises into linear factors in $E[x]$.

Solution

Elements of $F[x]/\langle h(x)\rangle$ have the form $a + b\alpha$ where $h(\alpha) = 0$, so that multiplication is carried out modulo $h(\alpha)$. Consider the polynomial ring $E[x]$ and the polynomial $h(x) \in E[x]$.

As $h(\alpha) = 0$, the remainder on dividing $h(x)$ by $x - \alpha$ is zero, that is, $h(x) = (x - \alpha)g(x)$ in $E[x]$. Now, $h(x)$ is a quadratic polynomial, and so $g(x)$ is linear and has a root $\beta \in E$. Thus both roots of $h(x)$ are in E and so E contains a splitting field for $h(x)$. Since any splitting field for $h(x)$ must contain F and α, and $F[x]/\langle h(x)\rangle \cong F(\alpha)$, $F(\alpha)$ is a splitting field for $h(x)$.

Worked Exercise 4.9

(a) Let F be a field. Show that for any non-zero $a \in F$, $x^2 + ax + a^2$ is irreducible in $F[x]$ if, and only if, $x^2 + x + 1$ is irreducible in $F[x]$.

(b) Let F be a field and let $a \in F$ with $a \neq 0$. Show that if $x^2 + x + 1$ is irreducible in F, then $F[x]/\langle x^2 + x + 1\rangle$ is a splitting field for $x^2 + ax + a^2$ over F.

Solution

(a) Suppose that $x^2 + x + 1$ is reducible in $F[x]$, so that there is $b \in F$ such that $b^2 + b + 1 = 0$. Then $x^2 + ax + a^2 = (x - ab)(x - ab^2)$, and so $x^2 + ax + a^2$ is reducible in $F[x]$.

Now suppose that $x^2 + ax + a^2$ is reducible in $F[x]$ so that there is a $c \in F$ with $c^2 + ac + a^2 = 0$. Letting $b = ca^{-1} \in F$ gives

$$\begin{aligned} b^2 + b + 1 &= (ca^{-1})^2 + ca^{-1} + 1 \\ &= c^2(a^{-1})^2 + ca(a^{-1})^2 + a^2(a^{-1})^2 \\ &= (c^2 + ac + a^2)(a^{-1})^2 = 0, \end{aligned}$$

so that b is a root of $x^2 + x + 1$ and $x^2 + x + 1$ is reducible in $F[x]$.

(b) By Theorem 2.19, we know that $F[x]/\langle x^2 + x + 1\rangle \cong F(\alpha)$, where α is such that $\alpha^2 + \alpha + 1 = 0$. The elements of $F(\alpha)$ have the form $u + v\alpha$ where $u, v \in F$, and $x^2 + x + 1$ is reducible in $F(\alpha)[x]$. Then, by part (a), $x^2 + ax + a^2$ is reducible in $F(\alpha)[x]$, and so $F(\alpha)$ contains a splitting field for $x^2 + ax + a^2$ over F.

Since $x^2 + x + 1$ is irreducible in $F[x]$, by part (a) $x^2 + ax + a^2$ is irreducible in $F[x]$, and so F is not a splitting field for $x^2 + ax + a^2$. As $[F(\alpha) : F] = 2$, there is no extension E of F with $F \subset E \subset F(\alpha)$, and so $F(\alpha) \cong F[x]/\langle x^2 + x + 1\rangle$ is a splitting field for $x^2 + ax + a^2$ over F.

After practising with quadratic polynomials, you may now wish to see a calculation of a splitting field for a cubic polynomial.

Example 4.10 *A splitting field for $x^3 - 2$ over $\mathbb{Q}$*

We are going to find a splitting field for $x^3 - 2$ over $\mathbb{Q}$. In Example 2.26 we found three extensions of $\mathbb{Q}$ isomorphic to $\mathbb{Q}[x]/\langle x^3 - 2\rangle$, one corresponding to each root of $x^3 - 2$ in $\mathbb{C}$.

Let $L = \mathbb{Q}(\sqrt[3]{2})$. We know that L is not a splitting field for $x^3 - 2$ since it does not contain the root $\omega\sqrt[3]{2}$, where $\omega = e^{\frac{2i\pi}{3}}$, as $L \subseteq \mathbb{R}$ and $\omega\sqrt[3]{2} \notin \mathbb{R}$.

Since we know that $\sqrt[3]{2}$ is a root of $x^3 - 2$ in $L[x]$, we can factorise this polynomial in $L[x]$ to obtain

$$x^3 - 2 = (x - \sqrt[3]{2})(x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^2).$$

Since $L = \mathbb{Q}(\sqrt[3]{2})$ is not a splitting field for $x^3 - 2$ over $\mathbb{Q}$, the quadratic $x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^2$ must be irreducible in $L[x]$ since otherwise it must factor into two linear factors, and this implies that both roots of the polynomial are in L.

By the solution to Worked Exercise 4.9, a splitting field over L for $x^2 + x + 1$ is also a splitting field for $x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^2$. It is therefore sufficient to consider the extension $L(\omega) \cong L[x]/\langle h(x)\rangle$ where $h(x) = x^2 + x + 1$ and ω is a root of $h(x)$.

Now as $h(\omega) = 0$, $h(x)$ is not irreducible in $L(\omega)[x]$. Hence by the solution to Worked Exercise 4.8, $L(\omega)$ is a splitting field for $h(x)$ over L. It is clear that $x^3 - 2$ splits over $L(\omega)$. We need to show that there is no field smaller than $L(\omega)$ over which $x^3 - 2$ splits.

We know that the elements of $L(\omega)$ can be written as $a + b\omega$ for $a, b \in L$ and the elements of L can be written as $r + s\alpha + t\alpha^2$ for $r, s, t \in \mathbb{Q}$ and $\alpha = \sqrt[3]{2}$. Thus, the elements of $L(\omega)$ can be written in the form

$$r + s\alpha + t\alpha^2 + (r_1 + s_1\alpha + t_1\alpha^2)\omega = r + s\alpha + t\alpha^2 + r_1\omega + s_1\alpha\omega + t_1\alpha^2\omega$$

where $r, s, t, r_1, s_1, t_1 \in \mathbb{Q}$.

Since α and $\alpha\omega$ must both be in a splitting field, ω must be in a splitting field, and since $\omega \notin L$, no field smaller than $L(\omega)$ can contain all these elements. Hence a splitting field for $x^3 - 2$ is $L(\omega) = \mathbb{Q}(\sqrt[3]{2})(\omega)$.

Note that by Theorem 2.29, the field $L(\omega) = \mathbb{Q}(\sqrt[3]{2})(\omega)$ is an extension of degree 6 over $\mathbb{Q}$. Equivalently, $L(\omega)$ is a vector space of dimension 6 over $\mathbb{Q}$.

In Example 4.10, we made considerable use of our knowledge of $\mathbb{C}$. Since the Fundamental Theorem of Algebra tells us that every non-constant polynomial over $\mathbb{C}$ factorises completely into linear factors, we know that any splitting field for a polynomial in $\mathbb{Q}[x]$ will be contained in $\mathbb{C}$ and will contain the roots of the polynomial. We may therefore refer to 'the' splitting field of a polynomial over $\mathbb{Q}$.

The crucial step in Example 4.10 was to observe that $x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^2$ is an irreducible polynomial over $L[x]$. The next example shows that this is not always the case in other fields.

Example 4.11 *A splitting field for $x^3 - 2$ over $\mathbb{Z}_{13}$*

Consider the polynomial $x^3 - 2 \in \mathbb{Z}_{13}[x]$. If a cubic polynomial is reducible over a field, it must have a linear factor and hence a root in that field. By inspection, we find that only $1, 5, 8, 12$ occur as cubes in $\mathbb{Z}_{13}$, and so $x^3 - 2$ is irreducible over $\mathbb{Z}_{13}$ since it has no root in $\mathbb{Z}_{13}$.

In order to determine the splitting field of $x^3 - 2$, we begin by forming the extension field $L = \mathbb{Z}_{13}[x]/\langle h(x)\rangle$, as in Example 4.10, and note that the elements of this field can be written in the form $r + s\alpha + t\alpha^2$ where $r, s, t \in \mathbb{Z}_{13}$ and $h(\alpha) = 0$, so that $\alpha^3 = 2$.

As before, $x^3 - 2 = (x - \alpha)(x^2 + \alpha x + \alpha^2)$, so we can refer to the solution of Worked Exercise 4.9 and note that $x^2 + \alpha x + \alpha^2$ is irreducible in $L[x]$ if, and only if, $x^2 + x + 1$ is irreducible in $L[x]$.

Since $x^3 - 1 = (x - 1)(x^2 + x + 1)$, the polynomial $x^2 + x + 1$ is reducible in $\mathbb{Z}_{13}[x]$ if, and only if, $\mathbb{Z}_{13}$ contains a primitive cube root of unity. Since $3^3 \equiv 1 \pmod{13}$, the field $\mathbb{Z}_{13}$ does contain a primitive cube root of unity, and so $x^2 + x + 1$ is reducible in $\mathbb{Z}_{13}[x]$. In fact, one may check that

$$x^2 + x + 1 = (x - 3)(x - 9) = (x + 10)(x + 4)$$

in $\mathbb{Z}_{13}[x]$, and so the polynomial is reducible in $L[x]$. Hence $x^2 + \alpha x + \alpha^2$ is reducible in $L[x]$ and so L contains a splitting field for $x^3 - 2$. Since any splitting field must contain α, the field L is a splitting field for $x^3 - 2$.

As an aside, note that the field $\mathbb{Z}_{13}(\sqrt[3]{2})$ has 13^3 elements. In the solution to Exercise 3.8 you saw a construction of a field with p^2 elements for any prime p such that $p \equiv 3 \pmod{4}$, and in Worked Exercise 1.8 you saw a field with 13^2 elements. This raises the natural question of what the number of elements of a finite field can be. We address this question in the next subsection and conclude with an exercise about splitting fields of cubic polynomials over finite fields.

Exercise 4.1

Find a splitting field for $x^3 - 2$ over

(a) $\mathbb{Z}_3$

(b) $\mathbb{Z}_5$.

4.2 Existence and uniqueness of finite fields

When we observed that we could regard a field extension as a vector space over the base field, we showed as a consequence that every finite field has order p^n for some n. In this last subsection of the chapter we will prove that not only does a field of order p^n exist for all primes p and all positive integers n but any two finite fields of the same order are isomorphic. This gives a classification of finite fields.

The approach to finding a field of order p^n is to find a polynomial of degree n and a field over which the polynomial splits. We then show that this field has order p^n.

Theorem 4.12

Let F be a finite field of order $q = p^n$ for some prime p. Then every element of F is a root of the polynomial $x^q - x$.

Proof The element $0 \in F$ is clearly a root of $x^q - x = x(x^{q-1} - 1)$. Thus, we only need to prove that any non-zero element of F is a root of the polynomial $x^{q-1} - 1$. As we know, the multiplicative group of F has order $q - 1$. By Lagrange's Theorem, the order of any element of a group is a divisor of the order of the group. Therefore for every x in the multiplicative group F^* $(= F\backslash\{0\})$ of the field F we have $x^{q-1} = 1$. Equivalently, every non-zero element of F is a root of $x^{q-1} - 1$. We conclude that every element of F is a root of the polynomial $x^q - x$. ■

We will gradually add more details to our setting for proving that for a given prime number p and any positive integer n there is a field of order p^n. As mentioned earlier, we will do this by constructing a splitting field of a specific polynomial over $\mathbb{Z}_p$. At some stage of the proof we will need to show that a certain polynomial has no repeated roots. We will do so by making use of an analogue of differentiation in a polynomial ring over a general field.

Definition 4.13 *Formal derivative*

Let K be a field. Let the mapping $d : K[x] \to K[x]$ be defined by

$$d(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_1 + 2a_2x + \cdots + na_nx^{n-1}.$$

We call the map d the **formal derivative**.

For polynomials over $\mathbb{R}$ or $\mathbb{C}$, the formal derivative is just the standard derivative.

The next lemma shows that the formal derivative behaves like the familiar derivative in real analysis as regards sums and products of polynomials.

Lemma 4.14

Let K be a field and let $d : K[x] \to K[x]$ be the formal derivative. Then, for any two polynomials $f(x), g(x) \in K[x]$, we have

(a) $d(f(x) + g(x)) = d(f(x)) + d(g(x))$

(b) $d(f(x)g(x)) = d(f(x))g(x) + f(x)d(g(x))$.

Proof

(a) Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ and $g(x) = b_0 + b_1x + \cdots + b_nx^n$. Then $f(x) + g(x) = a_0 + b_0 + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$, and

$$\begin{aligned} d(f(x) + g(x)) &= d(a_0 + b_0 + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n) \\ &= a_1 + b_1 + 2(a_2 + b_2)x + \cdots + n(a_n + b_n)x^{n-1} \\ &= a_1 + 2a_2x + \cdots + na_nx^{n-1} + b_1 + 2b_2x + \cdots + ba_nx^{n-1} \\ &= d(f(x)) + d(g(x)). \end{aligned}$$

(b) We prove the result in the special case $f(x) = ax^j$ and $g(x) = bx^k$. The general result then follows by part (a). We have

$$\begin{aligned} d(ax^jbx^k) &= d(abx^{j+k}) \\ &= (j + k)abx^{j+k-1} \\ &= jabx^{j+k-1} + kabx^{j+k-1} \\ &= jax^{j-1}bx^k + ax^j(kbx^{k-1}) \\ &= d(f(x))g(x) + f(x)d(g(x)). \end{aligned}$$

■

The formal derivative of Definition 4.13 is not the only map with properties (a) and (b) in Lemma 4.14; for example, the mapping $d_1 : K[x] \to K[x]$ defined by $d_1(f(x)) = xd(f(x))$ also satisfies them. A map that satisfy these two properties is called a **derivation** and it is a generalisation of the familiar operation of differentiation.

We are now ready to prove a lemma about the formal derivative that is an analogue of the familiar result from elementary calculus that a polynomial has a repeated root if, and only if, its derivative also has that root.

Lemma 4.15

Let K be a field and let $f(x)$ be a non-zero polynomial in $K[x]$. Let F be the splitting field for $f(x)$ over K. Then $f(x)$ has a root $\alpha \in F$ of multiplicity $m \geq 2$ if, and only if, the highest common factor of $f(x)$ and its formal derivative $d(f(x))$ is a polynomial of positive degree.

Proof Suppose that $\alpha \in F$ is a root of $f(x)$ of multiplicity $m \geq 2$. This is equivalent to saying that $f(x) = (x - \alpha)^m g(x)$ for some $g(x) \in F[x]$ and some $m \geq 2$.

The formal derivative d of $f(x)$ is given by

$$\begin{aligned} d(x^q - x) &= d((x-\alpha)^m g(x)) \\ &= m(x-\alpha)^{m-1} g(x) + (x-\alpha)^m d(g(x)) \\ &= (x-\alpha)^{m-1}[mg(x) + (x-\alpha)d(g(x))] \\ &= (x-\alpha)h(x), \end{aligned}$$

where $h(x) = (x-\alpha)^{m-2}[mg(x) + (x-\alpha)d(g(x))]$ is a (possibly constant) polynomial, since $m \geq 2$.

Thus the highest common factor of $f(x)$ and $d(f(x))$ is a multiple of $x - \alpha$, and so a polynomial of positive degree.

Now suppose that $f(x)$ and $d(f(x))$ are both multiples of a polynomial $h(x)$ that has positive degree. Then $h(x)$ has a root α in the splitting field F and so $f(x)$ and $d(f(x))$ are both multiples of $x - \alpha$. Let $f(x) = (x-\alpha)g(x)$. Then

$$\begin{aligned} d(f(x)) &= d((x-\alpha)g(x)) \\ &= (x-\alpha)d(g(x)) + g(x) \text{ by Lemma 4.14.} \end{aligned}$$

Since $d(f(x))$ is a multiple of $x - \alpha$, this implies that $g(x) = (x-\alpha)k(x)$ for some $k(x) \in F[x]$ and so $f(x) = (x-\alpha)^2 k(x)$. Thus $f(x)$ has a multiple root in the splitting field F. ■

We use this lemma to show that a particular polynomial has no repeated roots.

Lemma 4.16

Let $q = p^n$ be a prime power, let $f(x) = x^q - x \in \mathbb{Z}_p[x]$, and let F be a splitting field for $f(x)$ over $\mathbb{Z}_p$. Then the polynomial $f(x)$ has no repeated roots in F.

Proof By Corollary 4.5, there exists a splitting field for $f(x)$ over $\mathbb{Z}_p$. That is, there exists a field F containing $\mathbb{Z}_p$ such that every root of $f(x)$ lies in F, and (up to isomorphism) F is the smallest field with this property. Note that the characteristic of F is p since it contains $\mathbb{Z}_p$.

Suppose that $\alpha \in F$ is a root of $f(x)$ of multiplicity $m \geq 2$. Then, by Lemma 4.15, $d(f(x))$ is a multiple of $x - \alpha$, that is, $d(f(x)) = (x-\alpha)h(x)$ for some $h(x) \in F[x]$.

But $d(x^q - x) = qx^{q-1} - 1 = -1$ since $q \equiv 0 \pmod{p}$, which implies that $-1 = (x-\alpha)h(x)$. But this is absurd in $F[x]$. This contradiction shows that all roots of $f(x) = x^q - x$ in its splitting field F are distinct. ■

The next theorem is a significant result. It classifies all finite fields, and it is the aim we have been working towards with our discussion of splitting fields. A key step in the proof is to show that the splitting field of the polynomial $x^q - x$ for $q = p^n$ has order p^n.

Theorem 4.17 *Classification of finite fields*

For every prime p and every $n \geq 1$ there exists a field of order p^n, and any two fields of order p^n are isomorphic.

Proof *Existence* Let p and n be as in the statement, and let $q = p^n$. Consider the polynomial $f(x) = x^q - x$ over $\mathbb{Z}_p$. By Lemma 4.16 there exists a splitting field F for $f(x)$ over $\mathbb{Z}_p$ and $f(x)$ has no repeated roots in F.

Let A be the set of all the roots of $f(x)$ in the field F. Since the degree of $f(x)$ is q, the roots of $f(x)$ are mutually distinct, and since by the proof of Proposition 4.3 of Book C, Chapter 11, a polynomial of degree q with all roots distinct has exactly q roots in F, it follows that $|A| = q = p^n$. We would like to prove a perhaps unexpected fact – namely, that A is a subfield of F.

Firstly note that if $u \in A$ with $u \neq 0$ then $u^q - u = 0$ and so (using $\frac{1}{u}$ for u^{-1} for ease of reading) $\left(\frac{1}{u}\right)^q - \frac{1}{u} = \frac{u - u^q}{u^{q+1}} = 0$ and so $u^{-1} \in A$. Thus, to show that A is a subfield it is sufficient to convince ourselves that for any two elements $u, v \in A$ we have $u + v \in A$ and $uv \in A$; note that 0 and 1 belong to A since both are roots of $f(x)$.

Let $u, v \in A$, that is, $u^q - u = 0$ and $v^q - v = 0$, or, equivalently, $u^q = u$ and $v^q = v$. By Theorem 1.12 we have $(u + v)^q = u^q + v^q = u + v$. This means if u and v are roots of $f(x)$, then so is the element $u + v$. In other words, if $u, v \in A$, then $u + v \in A$. Moreover, $(uv)^q = u^q v^q = uv$. This shows that if $u, v \in A$, then $uv \in A$. It follows that A is a field, as claimed.

But since $A \subseteq F$ and F was assumed to be the smallest field containing all the roots of $f(x)$, we have $A = F$. As $|A| = |F| = q = p^n$, we have proved the existence of a field of order p^n.

Isomorphism Let K be a field of order $q = p^n$, and hence of characteristic p. By Theorem 1.11 we may assume that the field $\mathbb{Z}_p$ is a subfield of K. In the proof of Theorem 4.12 we saw that every element of K is a root of the polynomial $f(x) = x^q - x$ over $\mathbb{Z}_p$. Since this polynomial cannot have more than q roots, it follows that K is a splitting field for $f(x)$ over $\mathbb{Z}_p$.

Corollary 4.7 asserts that any two splitting fields of a polynomial over $\mathbb{Z}_p$ are isomorphic. It follows that any two finite fields of the same order are isomorphic. ■

Now that we know that a field of order p^n exists for every prime power, we show how to construct an irreducible polynomial of degree n over $\mathbb{Z}_p$.

Corollary 4.18

For any prime p and any positive integer n there exists an irreducible polynomial of degree n over $\mathbb{Z}_p$.

Proof By Theorem 4.17 there exists a field F of order p^n that contains $\mathbb{Z}_p$ as a subfield. Let ξ be a primitive element of F, that is an element that generates the multiplicative group $F\backslash\{0\}$.

Let us regard the field F as a vector space over $\mathbb{Z}_p$. The $n+1$ elements $1, \xi, \xi^2, \ldots, \xi^n$ must be linearly dependent, since the degree of F over $\mathbb{Z}_p$ is equal to n. It follows that there exist elements $a_0, a_1, \ldots, a_{n-1}, a_n \in \mathbb{Z}_p$ such that $a_0 1 + a_1\xi + \cdots + a_{n-1}\xi^{n-1} + a_n\xi^n = 0$. This means that ξ is a root of the polynomial $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n$ over $\mathbb{Z}_p$. We now show that $f(x)$ is irreducible and of degree n.

Suppose the contrary, that is, suppose that either $f(x)$ has degree smaller than n, or $f(x)$ is reducible over $\mathbb{Z}_p$. In either case we may assume that $f(x)$ has an irreducible divisor $h(x) \in \mathbb{Z}_p[x]$ of degree d, where $1 \leq d \leq n-1$, such that ξ is a root of $h(x)$.

Consider the field $E = \mathbb{Z}_p(\xi)$, that is, the smallest subfield of F containing $\mathbb{Z}_p$ and the element ξ. Since ξ is a root of an irreducible polynomial $h(x) \in \mathbb{Z}_p$ of degree $d < n$, by Proposition 2.19 all the elements of the field E have the form $b_0 + b_1\xi + \cdots + b_{d-1}\xi^{d-1}$, where $b_0, b_1, \ldots, b_{d-1} \in \mathbb{Z}_p$.

Since E has order p^d and contains the element ξ, the multiplicative order of ξ in E must be a divisor of $p^d - 1$. But E is a subfield of F and the orders of ξ in both these fields are the same. This contradicts our assumption that ξ is a primitive element, that is, an element of order $p^n - 1$, in F.

It follows that $f(x)$ is an irreducible polynomial of degree n over $\mathbb{Z}_p$, which completes the proof. ■

In the next worked exercise we will construct an irreducible polynomial of degree 2 over $\mathbb{Z}_3$ using the method outlined in the above proof. Of course, for degree 2 you would be able to find such a polynomial by other means, but finding one using the theorem in a simple case makes it easier to approach harder cases.

Worked Exercise 4.19

Use the approach given in the proof of Corollary 4.18 to find an irreducible polynomial of degree 2 over $\mathbb{Z}_3$.

Solution

The first step is to find a primitive element in the field of order 9. In order to do this we need to have a way of writing down the elements of the field. We use the solution to Exercise 3.3 where we found the primitive elements of the field $\mathbb{Z}_3[x]/\langle x^2+1\rangle = \{a + bi : a, b \in \mathbb{Z}_3\}$. (Yes, x^2+1 is irreducible over $\mathbb{Z}_3[x]$, but we will find another irreducible polynomial.) From Exercise 3.3 we know that $1 + i$ is a primitive element in this field.

The next step is to find $a, b, c \in \mathbb{Z}_3$, not all zero, such that

$$a + b(1 + i) + c(1 + i)^2 = 0$$

(a key point in the proof of Corollary 4.18 is that such a, b, c do exist). Since $i^2 = -1 = 2$ in $\mathbb{Z}_3$, this equation is equivalent to

$$(a + b) + (b + 2c)i = 0,$$

and, since 1 and i are linearly independent over $\mathbb{Z}_3$, it follows that $a + b = 0$ and $b + 2c = 0 \pmod 3$. This is a homogeneous linear system consisting of two equations and three unknowns and so there will be more than one solution. We, however, just need one (non-zero) solution. Letting, for example, $b = 1$ gives $c = 1$ and $a = 2$.

Thus, another example of an irreducible polynomial of degree 2 in $\mathbb{Z}_3[x]$ is $2 + x + x^2$.

This subsection, and this chapter, end with a worked exercise in which you are invited to follow the construction of an irreducible polynomial of degree 4 over $\mathbb{Z}_2$.

Worked Exercise 4.20

(a) Suppose that $\mathbb{Z}_2(\omega) \cong \mathbb{Z}_2[x]/\langle x^2 + x + 1\rangle$. Show that the polynomial $f(x) = x^2 + x + \omega$ is irreducible in $\mathbb{Z}_2(\omega)[x]$.

(b) Let $\mathbb{Z}_2(\omega, \alpha)$ be an extension of the field $\mathbb{Z}_2$ isomorphic to the field $\mathbb{Z}_2(\omega)[x]/\langle x^2 + x + \omega\rangle$, so that $\alpha^2 + \alpha + \omega = 0$. Show that $1 + \alpha$ is a primitive element of $\mathbb{Z}_2(\omega, \alpha)$.

(c) Use $1 + \alpha$ to find an irreducible polynomial of degree 4 over $\mathbb{Z}_2$.

Solution

(a) We have $f(0) = \omega$, $f(1) = \omega$, $f(\omega) = 1 + \omega$ and $f(1 + \omega) = 1 + \omega$. It follows that $f(x)$ has no linear factors and, since it is a quadratic polynomial, it is irreducible.

(b) The multiplicative group of $\mathbb{Z}_2(\omega, \alpha)$ has order $2^4 - 1$, that is, 15. Hence we need to show that $1 + \alpha$ has order 15. If it does not have order 15 it will have order 3 or 5, so we only need to check these powers of $1 + \alpha$. We will, however, calculate all the powers up to 5, as they will be needed in the final part of the solution. Since we are working modulo 2 and since $\omega^2 = \omega + 1$ and $\alpha^2 = \alpha + \omega$, we have

$$(1 + \alpha)^2 = 1 + \omega + \alpha$$
$$(1 + \alpha)^3 = 1 + \alpha + \omega\alpha$$
$$(1 + \alpha)^4 = \alpha$$
$$(1 + \alpha)^5 = \omega.$$

Since $1+\alpha$ does not have order 1, 3 or 5, it must have order 15 and it is, therefore, primitive.

(c) We wish to find $a, b, c, d, e \in \mathbb{Z}_2$, not all zero, such that

$$a + b(1+\alpha) + c(1+\alpha)^2 + d(1+\alpha)^3 + e(1+\alpha)^4 = 0.$$

Using the calculations in part (b) reduce this equation to

$$(a+b+c+d) + c\omega + (b+c+d+e)\alpha + d\omega\alpha = 0.$$

We know from the proof of Theorem 2.29 that the set $\{1, \omega, \alpha, \omega\alpha\}$ is linearly independent. Hence we deduce that $c = d = 0$ and so $a + b = 0$ and $b + e = 0$. There is just one non-zero solution, namely $a = b = e = 1$.

Thus, an irreducible polynomial of degree 4 over $\mathbb{Z}_2$ is $x^4 + x + 1$.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should be able to:

(a) Carry out calculations in various fields, including finite fields.

(b) Show given field extensions are isomorphic and show given field extensions are not isomorphic.

(c) Write down the basis for E as a field extension of a field F in simple cases.

(d) Find the minimal polynomial for an algebraic element over a field in simple cases.

(e) Find the inverse of elements in a given field extension.

(f) Determine whether a polynomial is irreducible over a given field in simple cases.

(g) Determine the splitting field of a polynomial over a field in simple cases.

(h) Determine whether a given element of $\mathbb{Z}_n$ and simple extensions is a primitive element.

Solutions to exercises

Solution to Exercise 1.1

We know from Proposition 4.14 of Chapter 17 that $\mathrm{Im}(f)$ is a ring. Since $\mathrm{Im}(f) \neq \{0\}$, by the solution to Exercise 1.4 of Book C, Chapter 11, the additive and multiplicative identities are distinct. To show that $\mathrm{Im}(f)$ is a field we must check that non-zero elements have multiplicative inverses in $\mathrm{Im}(f)$.

Let $r \in \mathrm{Im}(f)$ be such that $r \neq 0$. Then $r = f(s)$ for some $s \in F$, $s \neq 0$. Since F is a field, s has an inverse $s^{-1} \in F$. By Proposition 1.1 we then have $f(s^{-1}) = f(s)^{-1} = r^{-1}$ and hence $\mathrm{Im}(f)$ is a field.

Now $\mathrm{Ker}(f)$ is an ideal in F, and we know from the solution to Exercise 2.5 in Chapter 17 that the only ideals in F are $\{0\}$ or F itself. As $f \neq 0$, $\mathrm{Ker}(f)$ cannot be the whole field, so $\mathrm{Ker}(f) = \{0\}$. The First Isomorphism Theorem (Chapter 17, Theorem 4.17) gives $F/\mathrm{Ker}(f) \cong \mathrm{Im}(f)$. But since $\mathrm{Ker}(f) = \{0\}$, this implies that $F \cong \mathrm{Im}(f)$, as required.

Solution to Exercise 1.2

(a) Since there is no $\beta \in \mathbb{Z}$ with $p(\beta) = 0$, $p(x)$ has no linear factors and so it is irreducible by Gauss's Lemma.

(b) Although we could divide the product of two linear polynomials by $p(x)$ and find the remainder, we can instead use the fact that $\alpha^2 = \alpha - 1 \pmod{p(\alpha)}$. This is because

$$\begin{aligned}\alpha^2 &= x^2 + \langle p(x)\rangle \\ &= x^2 - x + 1 - (-x+1) + \langle p(x)\rangle \\ &= x - 1 + \langle p(x)\rangle.\end{aligned}$$

We can then calculate the product directly:

$$\begin{aligned}(a+b\alpha)(c+d\alpha) &= ac + (ad+bc)\alpha + bd\alpha^2 \\ &= ac + (ad+bc)\alpha + bd(\alpha - 1) \pmod{p(\alpha)} \\ &= ac - bd + (ad+bc+bd)\alpha \pmod{p(\alpha)}.\end{aligned}$$

Thus the multiplication rule for elements of E in the required representation is given by

$$(a+b\alpha)(c+d\alpha) = ac - bd + (ad+bc+bd)\alpha.$$

Solution to Exercise 1.3

(a) Since 5 and 6 are quadratic non-residues of 7, x^2-5 and x^2-6 are irreducible polynomials in $\mathbb{Z}_7[x]$. Hence $\mathbb{Z}_7[x]/\langle x^2-5\rangle$ and $\mathbb{Z}_7[x]/\langle x^2-6\rangle$ are fields by Proposition 1.5.

(b) By Proposition 1.7, the elements of $\mathbb{Z}_7[x]/\langle x^2-5\rangle$ can be written as $a+b\sqrt{5}$ where $a,b\in\mathbb{Z}_7$ and those of $\mathbb{Z}_7[x]/\langle x^2-6\rangle$ can be written as $c+d\sqrt{6}$ where $c,d\in\mathbb{Z}_7$.

We wish to define $\psi(a+b\sqrt{5})$ so that $\psi:\mathbb{Z}_7[x]/\langle x^2-5\rangle\to\mathbb{Z}_7[x]/\langle x^2-6\rangle$ is an isomorphism. In other words, we need to define $\psi(\sqrt{5})=r+s\sqrt{6}$ where $(r+s\sqrt{6})^2\equiv 5\pmod 7$.

Thus we require r and s such that

$$(r+s\sqrt{6})^2=r^2+6s^2+2rs\sqrt{6}\equiv 5 \pmod 7.$$

This implies that either r or s is zero. Since we know that 5 is a quadratic non-residue of 7, we have $r=0$. Thus we require s with $6s^2\equiv 5\pmod 7$, that is, such that $s^2\equiv 2\pmod 7$. Now 2 is a quadratic residue of 7 with $3^2\equiv 2\pmod 7$, so we can choose $s=3$.

Let $\psi(a+b\sqrt{5})=a+3b\sqrt{6}$ so that $\psi^{-1}(c+d\sqrt{6})=c+5d\sqrt{5}$, as $3\times 5\equiv 1\pmod 7$. Hence ψ is a bijection, and since $\psi(1)=1$ we just need to show the isomorphism properties for addition and multiplication.

For addition we have

$$\begin{aligned}\psi(a+b\sqrt{5})+\psi(c+d\sqrt{5}) &= a+3b\sqrt{6}+c+3d\sqrt{6}\\ &= (a+c)+3(b+d)\sqrt{6}\\ &= \psi(a+c+(b+d)\sqrt{5})\\ &\equiv \psi((a+b\sqrt{5})+(c+d\sqrt{5})) \pmod 7.\end{aligned}$$

For multiplication we have

$$\begin{aligned}\psi(a+b\sqrt{5})\psi(c+d\sqrt{5}) &= (a+3b\sqrt{6})(c+3d\sqrt{6})\\ &\equiv ac+5bd+3(ad+bc)\sqrt{6} \pmod 7\end{aligned}$$

since $54\equiv 5\pmod 7$, and

$$\begin{aligned}\psi((a+b\sqrt{5})(c+d\sqrt{5})) &= \psi(ac+5bd+(ad+bc)\sqrt{5})\\ &\equiv ac+5bd+3(ad+bc)\sqrt{6} \pmod 7.\end{aligned}$$

Thus ψ is a field isomorphism.

Solution to Exercise 1.4

(a) The degree of $x-2$ is 1, so by Proposition 1.7 $\mathbb{Q}[x]/\langle x-2\rangle$ is a field whose elements can be written in the form a for $a\in\mathbb{Q}$. The result follows.

(b) By the Fundamental Theorem of Algebra, the only irreducible polynomials in $\mathbb{C}[x]$ are polynomials of the form $x-\alpha$ where $\alpha\in\mathbb{C}$. The result follows as in part (a).

Solution to Exercise 1.5

Let $f\colon \mathbb{Z}_p \to F$ be defined by

$$f(n) = \underbrace{1+1+\cdots+1}_{n \text{ terms}}.$$

By the properties of addition and multiplication in $\mathbb{Z}_p$ and in F, f is a homomorphism and $\mathrm{Im}(f) = K$. The required result follows since the image of a field is a field by Exercise 1.1.

Solution to Exercise 2.1

Since any $a + b\alpha \in E$ can be written as $a \cdot 1 + b \cdot \alpha$, $\{1, \alpha\}$ is a spanning set. Since the elements of E are uniquely expressed in the form $a + b\alpha$, where $a, b \in \mathbb{Q}$, $a + b\alpha = 0$ only when $a = b = 0$, and so the set $\{1, \alpha\}$ is linearly independent. Thus $\{1, \alpha\}$ is a basis for E as a vector space over F.

There are many other bases, but $\{1, \alpha\}$ is a rather natural one to consider.

Solution to Exercise 2.2

Let $m(x)$ be the minimal polynomial for α over F. Then $m(x)$ is a monic polynomial of lowest degree for which $m(\alpha) = 0$.

Therefore $\deg h(x) \geq \deg m(x)$, and so by the Division Algorithm for polynomials over a field we can write

$$h(x) = q(x)m(x) + r(x),$$

where $\deg r(x) < \deg m(x)$. Since $h(\alpha) = m(\alpha) = 0$ we have that $r(\alpha) = 0$. But as $m(x)$ is the smallest non-zero polynomial satisfying $m(\alpha) = 0$, we conclude that $r(x) = 0$ and so $h(x) = q(x)m(x)$. As $h(x)$ is monic and irreducible, this is only possible if $h(x) = m(x)$.

Solution to Exercise 2.3

(a) $x^2 + 1$

(b) (i) Since neither 2 nor 3 is a quadratic residue in $\mathbb{Z}_5$, the minimal polynomials are $x^2 - 2$ and $x^2 - 3$ respectively.

(ii) In $\mathbb{Z}_5(\beta)$ we have that $(2\beta)^2 = 4\beta^2 = 2 \pmod 5$ and $(3\beta)^2 = 9\beta^2 = 2 \pmod 5$, so $\alpha \in \mathbb{Z}_5(\beta)$ and $\mathbb{Z}_5(\alpha) \subseteq \mathbb{Z}_5(\beta)$. Since both extensions have degree 2 over $\mathbb{Z}_5$ by Theorem 2.21(c), a basis for $\mathbb{Z}_5(\alpha)$ is also a basis for $\mathbb{Z}_5(\beta)$, and so the two extensions must in fact be equal.

Solution to Exercise 2.4

As neither 1 nor -1 are roots of the equation $x^3 - 3x - 1 = 0$, by the Rational Root Test (Book C, Chapter 11, Theorem 4.7), the polynomial has no linear factors and so it is irreducible over $\mathbb{Q}$. As it is monic, it is the minimal polynomial for α. Since it has degree 3, the degree of the extension $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$ is 3 by Theorem 2.21(c).

Solution to Exercise 2.5

Let $m(x)$ be the polynomial obtained by dividing $h(x)$ through by the leading coefficient. Then $m(x)$ is a monic and irreducible polynomial with $m(\alpha) = m(\beta) = 0$. Thus $m(x)$ is the minimal polynomial for both α and β by the solution to Exercise 2.2. The required result follows by Theorem 2.21.

Alternatively, the result follows from a direct allocation of Corollary 2.23.

Solution to Exercise 2.6

The solution to Worked Exercise 1.8 showed that $\mathbb{Z}_{13}(\sqrt{2}) \cong \mathbb{Z}_{13}(\sqrt{5})$. It is not difficult to see that the map from $\mathbb{Z}_{13}(\sqrt{t})$ to $\mathbb{Z}_{13}(\sqrt{-t})$ defined by $a + b\sqrt{t} \mapsto a + b\sqrt{-t}$ is an isomorphism, so we only need to show that $\mathbb{Z}_{13}(\sqrt{2}) \cong \mathbb{Z}_{13}(\sqrt{6})$. We use the same approach as in Worked Exercise 1.8 to work out the image of $\sqrt{2}$ under an isomorphism $\phi : \mathbb{Z}_{13}(\sqrt{2}) \mapsto \mathbb{Z}_{13}(\sqrt{6})$.

Let $\phi(\sqrt{2}) = a + b\sqrt{6}$. Then $2 = (a + b\sqrt{6})^2 = a^2 + 6b^2 + 2ab\sqrt{6}$, so $a = 0$ or $b = 0$. If $a = 0, 6b^2 = 2$ and so $3b^2 = 1$ and so $b^2 = 9$ and $b = 3$ (or 10). Thus we let $\phi(\sqrt{2}) = 3\sqrt{6}$, and so $\phi(a + b\sqrt{2}) = a + 3b\sqrt{6}$. The morphism property for addition is clear. We just need to check the property for multiplication:

$$\begin{aligned}\phi(a + b\sqrt{2})\phi(c + d\sqrt{2}) &= (a + 3b\sqrt{6})(c + 3d\sqrt{6}) \\ &= (ac + 2bd) + (3(bc + ad)\sqrt{6}) \\ &= \phi(ac + 2bd + (bc + ad)\sqrt{2}) \\ &= \phi[(a + b\sqrt{2})(c + d\sqrt{2})].\end{aligned}$$

Since the homomorphism property holds for both addition and multiplication, ϕ is an isomorphism, as required.

Solution to Exercise 2.7

The quadratic formula shows that $j = \dfrac{2 \pm \sqrt{-4}}{2} = 1 \pm i$ and so $j \in \mathbb{Q}(i)$. Thus $\mathbb{Q}(j) \subseteq \mathbb{Q}(i)$. Since $[\mathbb{Q}(i) : \mathbb{Q}] = [\mathbb{Q}(j) : \mathbb{Q}] = 2$, this is only possible if $\mathbb{Q}(j) = \mathbb{Q}(i)$.

Solution to Exercise 2.8

Suppose, by way of contradiction, that $\phi : \mathbb{Q}(\xi) \mapsto \mathbb{Q}(\eta)$ is an isomorphism sending ξ to $a + b\eta$ for some $a, b \in \mathbb{Q}$. Since both fields are subfields of $\mathbb{C}$, we know from Theorem 1.14 that $\phi(x) = x$ for all $x \in \mathbb{Q}$. In particular $\phi(-2) = -2$ so that

$$-2 = \phi(\xi^2) = (\phi(\xi))^2 = a^2 + 2ab\eta + b^2\eta^2 = a^2 + 2b^2 + 2b(a + b)\eta,$$

since $\eta^2 = 2\eta + 2$.

Since $\{1, \eta\}$ is a basis for $\mathbb{Q}(\eta)$, we can compare coefficients and deduce that $a^2 + 2b^2 = -2$, which is impossible for $a, b \in \mathbb{Q}$. So the fields $\mathbb{Q}(\xi)$ and $\mathbb{Q}(\eta)$ are not isomorphic.

Solution to Exercise 2.9

(a) We have $\mathbb{Q}(\sqrt{2}) = \mathbb{Q}(-\sqrt{2})$, since $-\sqrt{2} = (-1) \times \sqrt{2} \in \mathbb{Q}(\sqrt{2})$ and $\sqrt{2} = (-1) \times (-\sqrt{2}) \in \mathbb{Q}(-\sqrt{2})$.

(b) Elements of $\mathbb{Q}(\sqrt{-2})$ have the form $a + bj$ where $j^2 = -2$. Suppose that there is an isomorphism $\phi\colon \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{-2})$. Then $\phi(\sqrt{2}) = a + bj$ for some $a, b \in \mathbb{Q}$. Since $\phi(2) = 2$ we have

$$2 = \phi(2) = \phi(\sqrt{2})\phi(\sqrt{2}) = (a + bj)^2 = a^2 - 2b^2 + 2abj.$$

From this we can see that $a = 0$ or $b = 0$. If $b = 0$ then $a^2 = 2$, which is not possible for $a \in \mathbb{Q}$. If $a = 0$ then $-2b^2 = 2$, which is not possible, either, for $b \in \mathbb{Q}$. Hence there can be no isomorphism ϕ and $\mathbb{Q}(\sqrt{2}) \ncong \mathbb{Q}(\sqrt{-2})$.

Solution to Exercise 2.10

Since $(a + b\sqrt{2})^2 = a^2 + 2b^2 + 2ab\sqrt{2}$, if $(a + b\sqrt{2})^2 = 3$ then either $a = 0$ or $b = 0$. If $a^2 = 0$ then $b^2 = \frac{3}{2}$ and so $(2b)^2 = 6$, which is not possible for $b \in \mathbb{Q}$. If $b^2 = 0$ then $a^2 = 3$, which is not possible, either, for $a \in \mathbb{Q}$.

Thus, there are no $a, b \in \mathbb{Q}$ such that $(a + b\sqrt{2})^2 = 3$. Hence $x^2 - 3$ is irreducible over L. This shows that $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}(\sqrt{2})] = 2$. Since $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$, Theorem 2.29 implies that $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$.

Solution to Exercise 3.1

(a) For any $u^j \in U$ we have $(u^j)^d = (u^d)^j = 1$, and so u^j is a root of $x^d - 1$.

(b) If $u^i = u^j$, then $u^{j-i} = 1$, which is impossible since $0 < j - i < d$ and u has order d.

Solution to Exercise 3.2

Let $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ and suppose that $n = rd$. Then r has order d and $\{r, 2r, \ldots, (d-1)r\}$ are the only elements whose order divides d. If s and d have a common factor t $(\neq 1)$, say, then the order of sr divides $\frac{d}{t}$, since $\frac{sd}{t}$ is a multiple of d. Otherwise sr has order d. Thus, all elements of order exactly d in the additive group $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ have the form sr, where $n = rd$, $1 \leq s \leq d$ and s and d are relatively prime. Hence the number of elements of order d in $(\mathbb{Z}_n, +)$ is equal to $\phi(d)$.

Solution to Exercise 3.3

(a) $(1 + i)$, $(1 + i)^2 = 2i$, $(1 + i)^3 = 1 + 2i$, $(1 + i)^4 = 2$, $(1 + i)^5 = 2 + 2i$, $(1 + i)^6 = i$, $(1 + i)^7 = 2 + i$, $(1 + i)^8 = 1$. (Remember that here arithmetic is modulo 3.)

(b) By part (a), $1 + i$ is a primitive element. Moreover, so are $(1 + i)^3 = 1 - i = 1 + 2i$, $(1 + i)^5 = 2 + 2i$ and $(1 + i)^7 = 2 + i$, as $3, 5$ and 7 are all coprime to 8.

Solution to Exercise 3.4

The possible orders of an element of $\mathbb{Z}_5^*$ are 1, 2 and 4, so it is enough to check that $2^2 \not\equiv 1 \pmod 5$. In fact $2^2 \equiv 4 \pmod 5$, so 2 is a primitive element modulo 5.

The possible orders of an element of $\mathbb{Z}_{11}^*$ are 1, 2, 5 and 10, so it is enough to check that $2^2 \not\equiv 1 \pmod{11}$ and $2^5 \not\equiv 1 \pmod{11}$. In fact $2^2 \equiv 4 \pmod{11}$ and $2^5 \equiv 10 \pmod{11}$, so 2 is a primitive element modulo 11.

The possible orders of an element of $\mathbb{Z}_{13}^*$ are 1, 2, 3, 4, 6 and 12, so it is enough to check that $2^2, 2^3, 2^4, 2^6 \not\equiv 1 \pmod{13}$. Modulo 13 we have

$$2^2 \equiv 4,\ 2^3 \equiv 8,\ 2^4 \equiv 3,\ 2^6 \equiv 2^2 \cdot 2^4 \equiv 12,$$

so 2 is a primitive element modulo 13.

The possible orders of an element of $\mathbb{Z}_{19}^*$ are 1, 2, 3, 6, 9 and 18, so it is enough to check that $2^2, 2^3, 2^6, 2^9 \not\equiv 1 \pmod{19}$. Modulo 19 we have

$$2^2 \equiv 4,\ 2^3 \equiv 8,\ 2^6 \equiv 8^2 \equiv 7,\ 2^9 \equiv 2^3 \cdot 2^6 \equiv 8 \cdot 7 \equiv 18$$

so 2 is a primitive element modulo 19.

Solution to Exercise 3.5

The possible orders of an element of $\mathbb{Z}_7^*$ are 1, 2, 3 and 6, so it is enough to check that $3^2 \not\equiv 1 \pmod 7$ and $3^3 \not\equiv 1 \pmod 7$. In fact $3^2 \equiv 2 \pmod 7$ and $3^3 \equiv 6 \pmod 7$, so 3 is a primitive element modulo 7.

The possible orders of an element of $\mathbb{Z}_{17}^*$ are 1, 2, 4, 8 and 16, so it is enough to check that $3^8 \not\equiv 1 \pmod{17}$ (as this also covers the powers 2 and 4). In fact, $3^8 \equiv 16 \pmod{17}$, so 3 is a primitive element modulo 17.

The possible orders of an element of $\mathbb{Z}_{19}^*$ are 1, 2, 3, 6, 9 and 18, so it is enough to check that $3^6 \not\equiv 1 \pmod{19}$ and $3^9 \not\equiv 1 \pmod{19}$ (as these also cover the powers 2, 3 and 6). In fact $3^6 \equiv 7 \pmod{19}$ and $3^9 \equiv 18 \pmod{19}$, so 3 is a primitive element modulo 19.

Solution to Exercise 3.6

$2^n \pmod{11}$	1	2	3	4	5	6	7	8	9	10
n	10	1	8	2	4	9	7	3	6	5

(a) The non-zero elements of $\mathbb{Z}_{11}$ that are squares are $1, 3, 4, 5, 9$, and so the values of a for which $t^2 - a$ is irreducible will be $2, 6, 7, 8, 10$.

(b) All elements of $\mathbb{Z}_{11}$ are cubes, and so all polynomials $t^3 - a$ are reducible over $\mathbb{Z}_{11}$.

(c) The primitive elements are $2 = 2^1$, $6 = 2^9$, $7 = 2^7$, $8 = 2^3$. (These are the elements whose exponent as a power of 2 is coprime to 10.)

Solution to Exercise 3.7

$3^n \pmod{17}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
n	16	14	1	12	5	15	11	10	2	3	7	13	4	9	6	8

(a) The non-zero elements of $\mathbb{Z}_{17}$ that are squares are $1, 2, 4, 8, 9, 13, 15, 16$, and so the values of a for which $t^2 - a$ is irreducible are $3, 5, 6, 7, 10, 11, 12,$ 14.

(b) All elements of $\mathbb{Z}_{17}$ are cubes, and so all polynomials $t^3 - a$ (for $a \in \mathbb{Z}_{17}^*$) are reducible over $\mathbb{Z}_{17}$.

(c) The primitive elements are $3 = 3^1$, $5 = 3^5$, $6 = 3^{15}$, $7 = 3^{11}$, $10 = 3^3$, $11 = 3^7$, $12 = 3^{13}$, $14 = 3^9$. (These are the elements whose exponent as a power of 3 is coprime to 16.)

Solution to Exercise 3.8

By Proposition 3.9, $x^2 + 1$ is irreducible in $\mathbb{Z}_p[x]$. Thus by Theorem 2.19, the field $\mathbb{Z}_p[x]/\langle x^2 + 1\rangle$ can be regarded as an extension of $\mathbb{Z}_p$ containing p^2 elements that have the form $a + b\alpha$ where $a, b \in \mathbb{Z}_p$ and $\alpha^2 + 1 = 0$.

Solution to Exercise 4.1

(a) In $\mathbb{Z}_3$, $x^3 - 2 = (x - 2)^3$ (by Theorem 1.12), and so all the roots of $x^3 - 2$ are in $\mathbb{Z}_3$ and the splitting field is $\mathbb{Z}_3$ itself.

(b) In $\mathbb{Z}_5$, $x^3 - 2 = (x - 3)(x^2 + 3x + 4)$. Now $x^2 + 3x + 4$ has no roots in $\mathbb{Z}_5$, and so it is irreducible in $\mathbb{Z}_5[x]$. Thus, a splitting field is $\mathbb{Z}_5[x]/\langle x^2 + 3x + 4\rangle$.